Centrale Nantes

UNIVERSITE DE NANTES

MASTER SCIENCES MECANIQUES APPLIQUEES

SPECIALITE GENIE MECANIQUE

Année 2006/2007

Diplôme cohabilité par

l'École Centrale Nantes et l'Université de Nantes

Mémoire de master

Présenté et soutenu par:

PHILIPPE RAUFFET

le 10 octobre 2007

à l'Ecole Centrale Nantes

TITRE

A METHODOLOGY FOR THE APPLICATION OF AN AUTOMATED AND INTERACTIVE REIFICATION PROCESS IN A VIRTUAL COMMUNITY OF PRACTICE

JURY

ALAIN BERNARD

CATHERINE DA CUNHA

MICHEL LABROUSSE

« La connaissance n'est pas réservée aux seuls vieillards : un enfant qui sait interroger peut l'acquérir ».
(Massa Makan Diabaté)

«Men can know more than their ancestors did if they start with a knowledge of what their ancestors had already learned....That is why a society can be progressive only if it conserves its traditions. » (Walter Lippmann)

«Knowledge is of two kinds. We know a subject ourselves, or we know where we can find information on it. »
(Samuel Johnson)

AVANT-PROPOS

Ce mémoire s'inscrit dans le cadre du master de Sciences mécaniques Appliquées, diplôme cohabilité par l'école Centrale Nantes et l'Université de Nantes.

Il vise à transcrire mes travaux de recherches, qui se sont déroulés au sein de l'entreprise Indutech, à Stellenbosch en Afrique du Sud. Ceux-ci ont porté essentiellement sur les communautés de pratique et leur évolution, l'analyse de contenu documentaire, la représentation des concepts et le filtrage collaboratif.

Il fait suite à un travail préliminaire de mémoire bibliographique, soutenu le 30 mars 2007 à l'école Centrale Nantes, dans lequel sont rappelés les principales définitions, les méthodes et les outils, relatifs à la connaissance et à la gestion des connaissances.

FOREWORD

This report occurs within the frame of master Applied Mechanical Science, diploma delivered by Centrale Nantes and Nantes University.

It aims at transcribing my research works, which took place in the firm Indutech, in Stellenbosch, South Africa. This one deals mainly with the Communities of practice and their evolution, document content analysis, concepts mapping and collaborative filtering.

It follows a preliminary work of literature study, presented on March, 30th 2007 in Centrale Nantes, giving the main definitions, methods and tools, in relation with knowledge and knowledge management.

RESUME EN FRANCAIS

Cadre général

Dans un contexte industriel en perpétuelle mutation, les entreprises doivent faire face à des changements continus et incertains qui les obligent à innover et à se moderniser en permanence. Dans un souci de flexibilité, elles doivent de plus se montrer aptes à fournir des solutions rapides et adaptées aux besoins de leurs clients.

Pour cela, les entreprises se sont d'abord restructurées, en externalisant et en réduisant la taille de leurs organisations, ce qui a provoqué une perte de connaissance. Cependant, elles considèrent aujourd'hui que la mondialisation entraîne un besoin croissant pour le partage du capital immatériel. En réponse à ce constat, elles ont créé des groupes transversaux et flexibles pour échanger et innover, qui sont appelées Communautés de Pratiques (CoPs).

Vers une problématique : étude des CoPs et de leurs limites dans un environnement mondialisé, virtuel et numérique

Dans la première partie de ce mémoire, nous avons étudié et défini les CoPs au travers de deux caractéristiques fondamentales :

- Le principe de Participation Périphérique Légitimée, qui décrit le processus d'apprentissage contextualisé au sein des CoPs,
- La dualité réification/participation, qui discrimine les actions des membres des CoP en deux catégories, les tâches de génération d'information d'une part, et le travail collaboratif et le partage d'information d'autre part.

Le cycle de vie de la connaissance a aussi été analysé afin de mesurer les impacts des CoPs sur celui-ci.

Nous avons ensuite observé les Communautés de Pratiques dans leur environnement actuel et « globalisé », ainsi que leurs changements dans ce nouveau cadre. D'abord locales et contextualisées, les CoPs ont en effet du s'adapter à de profondes transformations. Les progrès en numérisation et l'apparition des communautés internet ont modifié le fonctionnement de ces réseaux de

connaissances, les rendant virtuels et les organisant autour de bases de connaissance numériques.

Ces changements créent de nouvelles barrières et limites. La masse d'information rend l'organisation des bases de connaissances complexes, et la virtualité des réseaux affectent les deux caractéristiques fondamentales des CoPs:

- Le principe de Participation Périphérique Légitimée est altérée (en raison d'un nombre réduit de communication face à face, ainsi que du manque de temps pour apprendre des autres et évaluer leurs expertises et leurs légitimités)
- La dualité participation/réification est déséquilibrée (la part concernant la génération d'information prédomine sur celle des échanges et des discussions au sein des nouvelles CoPs virtuelles).

A partir de ces remarques, la définition de la problématique du mémoire a été guidée par le souci de dépasser ces barrières. Limitant la réflexion au cadre de la réification (i.e. la manière de générer l'information et de la pousser aux membres des CoPs), le problème a été approché et décomposé en trois sous-problèmes, suivant les aspects mis en évidence dans le principe de Participation Périphérique Légitimée.

Ont été ainsi soulevées les questions de l'automatisation de la réification, de l'enrichissement et du contrôle de cette information extraite automatiquement, et de l'évaluation des membres des CoPs et de leurs actions.

<u>Etat de l'art : techniques et méthodes pour automatiser la génération d'information, et contrôler et enrichir le processus d'automatisation</u>

La problématique a ensuite orienté l'état de l'art de la deuxième partie, lequel a été mené pour trouver des réponses techniques et méthodologiques aux questions posées.

Nous avons étudié les différents outils et méthodes pour l'automatisation de la réification, de l'extraction de l'information à sa visualisation, en passant par son organisation. Puis ont été analysées les possibilités pour l'utilisateur de compléter et d'évaluer le contenu extrait, ainsi que les méthodes informatiques pour enregistrer et déduire d'autres informations des interactions de l'utilisateur avec le système.

<u>Développement : proposition d'une méthodologie et études de cas</u>

Combinant la problématique et les observations de l'état de l'art, nous avons enfin tenté de développer une méthodologie dans une troisième et dernière partie, afin d'appliquer une réification automatisée et interactive au sein de Communautés de Pratiques virtuelles.

Après la définition et la spécification des besoins, les différentes étapes de la méthodologie ont été proposées, expliquées par des modèles et des exemples.

Afin d'accompagner cette méthodologie et dans la perspective de l'implémenter, nous avons choisi les outils les plus adaptés parmi ceux passés en revue dans l'état de l'art.

Finalement, nous avons réalisé deux études de cas, mettant en œuvre et plaçant dans un contexte plus pragmatique la méthodologie proposée :

- Les outils d'Indutech, l'entreprise spécialisée en gestion de l'innovation dans laquelle j'ai effectué mon stage de recherches, ont été testés et analysés pour montrer comment ils pouvaient supporter la méthodologie,
- Une étude de cas théorique a été conduite sur une CoP virtuelle européenne, VRL-KCiP, afin de commenter l'apport de la méthodologie en situation.

Table of Contents

INTRODUCTION 1

PART A. OBSERVATIONS & PROBLEMS 5

Chapter I. Communities of Practice & Knowledge background 7

I.1. Preliminary background: towards a knowledge life cycle 7

I.1.1. Between epistemology and systemic approach 7

I.1.2. Typologies of knowledge and artefacts 8

I.1.3. Knowledge life cycle 9

I.2. Communities of Practice: conceptual presentation 12

I.2.1. Definition 12

I.2.2. Typologies 12

I.2.3. Roles of CoPs 13

I.2.4. Needs of CoPs 14

I.3. Two Wenger's models and knowledge life cycle in CoPs 14

II.3.1. Legitimate Peripheral Participation 14

II.3.2. Participation-reification duality and modified knowledge life cycle 17

I.3. Conclusion 20

Chapter II. Evolution of CoPs, barriers & limits 21

II.1. CoPs, technological progresses and questions 21

II.1.1. Virtualisation and growing importance of artefacts 21

II.1.2. artefact digitalization 22

II.1.3. Social networks, blogs and wikis 23

II.2. Barriers and limits of virtual and digital CoPs 24

II.2.1. Barriers: new CoPs and LPP Process 24

II.2.2. Limits: socialization versus internalization in learning process 25

II.3. Conclusion 26

Chapter III. Hypothesis & Thesis Problems 27

III.1. Hypothesis, framework of thesis problems 27

III.2. Thesis problems 28

III.3. Conclusion 29

PART B. STATE OF ART 30

Section 1: Computerized reification 32

Chapter IV. Document content analysis 33

IV.1. Introduction to content analysis 33

IV.1.1. Definition 33

IV.1.2. Goals 34

IV.1.3. Requirements 34

IV.1.4. Limits 34

IV.1.5. Typologies 35

IV.2. Text analysis methods 37

IV.2.1. Text analysis in content analysis 37

IV.2.3. Linguistic and statistical method: Usual Text Mining 38

IV.2.3. Non linguistic and unsupervised method: Topic Modelling 40

IV.3. Conclusion 42

Chapter V. Information architecture with “knowledge maps”..... 43

V.1. Introduction to knowledge maps 43

V.1.1. Definition 43

V.1.2. Classification 44

V.2. Knowledge map structures 44

V.2.1. Conceptual maps, an informal means to organize concepts 44

V.2.2. Ontology, a formal architecture 45

V.2.3. Topic maps, semi-formalized tools 47

V.3. Conclusion 49

Chapter VI. Visualisation tools 51

VI.1. Interest and roles of information visualization 51

VI.2. Visualisation tools 52

VI.2.1.Graph and Tree tools 53

VI.2.2. Map tools 54

VI.2.3. Map and tree combinations 55

VI.3. Comparisons of tools 56

VI.4. Map automation 58

VI.5. Conclusion 58

Section 2: Enrichment & assessment 59

Chapter VII. The link between manual enrichment & assessment60

VII.1. Enrichment and control of information by users 60

VII.1.1. Roles and needs 60

VII.1.2. Actions 61

VII.1.3. Existing solutions 62

VII.2. Assessment 64

VII.2.1. Performance 64

VII.2.2. Indicators of performance 65

VII.2.3. Assessment in the context of reification and virtual CoPs 66

VII.3. Enrichment and evaluation, two complementary tasks 68

VII.4. Conclusion 69

Chapter VIII. Information filtering 70

VIII.1. Recommender systems 70

VIII.1.1. Definition 70

VIII.1.2. Two approaches, two benchmarks 71

VIII.2. Information filtering methods 71

VIII.2.2. “Individual filtering” and “user profiling” 71

VII.2.3. Social or collaborative filtering 73

VII.2.4. Example 74

VIII.3. Information filtering for enrichment and evaluation 75

VIII.4. Conclusion 76

PART C: DEVELOPMENT 77

Chapter IX. Proposed methodology 79

IX.1. Specifications 79

IX.1.1. Definitions of Goals 79

IX.1.2. “Book of specifications” 80

IX.2. Analysis of specifications 82

IX.2.1. Scenario-based analysis 82

IX.2.2. Identification of involved items 84

IX.3. Methodology and tools 85

IX.3.1. Proposition of a methodology 85

IX.3.2. Summary and models for the proposed methodology 90

IX.4. Tools & Methods 94
IX.4.1. Suggestions for tools and methods supporting the methodology 94
IX.4.2. Suggestion for a system of "indicators of performance" 97
IX.5. Conclusion 98

Chapter X. The methodology and Indutech's tools 100

X.1. CAT, a topic modelling software 100
X.1.1. Presentation 100
X.1.2. Tests and analysis of the efficiency and the effectiveness of CAT 101
X.1.3. Assets for supporting the methodology 105
X.2. Organon, a conceptual framework browser 105
X.2.1. Presentation 105
X.2.2. Assets for supporting the methodology 106
X.3. EDEN™, an Enterprise-wide Innovation Management Platform 106
X.3.1. Presentation 106
X.3.2. Assets for supporting the methodology 107
X.3. Conclusion 108

Chapter XI. Theoretical Case Study on VRL-KCiP 110

XI.1. VRL-KCiP, a virtual Community of Practice 110
XI.2. Theoretical study: impacts of the methodology on VRL's needs 112
XI.2.1. The limited scope of the study 113
XI.2.2. VRL's needs & specifications: Task 105 & Overview for the Associate members of the VRL 114
XI.2.3. Analysis of VRL's needs and specifications 116
XI.2.4. Responses of proposed methodology and tools for VRL 118
XI.3. Conclusion 120

CONCLUSIONS AND TRENDS 121

APPENDICES 128

REFERENCES 150

LIST OF FIGURES

Figure 1: Processes in systemic view (Judelman, 2004) 9
Figure 2: From data to knowledge and from knowledge to information (Zimmermann, et al., 2002) .. 9
Figure 3: SECI Model (Nonaka, et al., 2000) 10
Figure 4: proposition of a divided knowledge life cycle combining SECI view and systemic view 11
Figure 5: Division by pairs of LPP model 16
Figure 6: Participation/reification duality (Wenger, 1998) 18
Figure 7: Divided SECI model with the Wenger's Duality 19
Figure 8: participation in virtual environment (Li, 2007) 25
Figure 9: Summary of thesis problem 29
Figure 10: Tree of Porphyry drawn by Peter of Spain (1329) 32
Figure 11: Text analysis methods in content analysis 37
Figure 12: basic text mining process (Even-Zohar, 2002) 38
Figure 13: roles and power of visualization (Judelman, 2004) 52
Figure 14: Example of graphs (Netvis, 2007) 53
Figure 15: example of tree (contemplativemind, 2007) 53
Figure 16: Example of map using topography, colour, size and distance (Dodge, 2000) 54
Figure 17: Example of cluster map on the insurance world (TouchGraph, 2007) 55
Figure 18: Example of fractal map (Grokker, 2007) 55
Figure 19: Example of Tree map on television choices (The Hive Group, 2007) 56
Figure 20: Feedback and control in automation (Edinburgh University, 2007) 59
Figure 21: del.icio.us, a folksonomy website 62
Figure 22: adding keywords with EDEN 63
Figure 23: adding a summary describing the content of a document with EDEN 63
Figure 24: ratings system of Youtube 63
Figure 25: Tetrahedron of performance (Bescos, et al., 1995) 64
Figure 26: DeLone and McLean's IS success model (Delone, et al., 1992) 66
Figure 27: Link between manual enrichment and assessment 68
Figure 28: Example of information filtering on YouTube 74
Figure 29: Scenarios of user's activities 82
Figure 30: SADT, links between reification, interaction, evaluation and enrichment 92
Figure 31: Flow based-model 93
Figure 32: Influence of the size of corpus 102
Figure 33: influence of the number of topics 102

Figure 34: Combined influence of S and K 103
Figure 35: Time of processing ~N L $K^{0.5}$ 103
Figure 36: Combined use of CAT, Organon and EDEN 108
Figure 37: VRL's Logo (VRL-KCiP, 2007) 110
Figure 38: VRL partners 111
Figure 39: Limits of the study 113
Figure 40: Limits of the case study 114
Figure 41: Analysis of VRL's needs and specifications 117
Figure 42: Methodology and toolbox versus VRL's needs and specifications 119
Figure 43: overview on the work done 122
Figure 44: Summary of the thesis problem 123
Figure 45: Summary of the State of Art 124
Figure 46: Summary of the development 125

LIST OF TABLES

Table 1: evolution of knowledge transmission and learning (Judelman, 2004) 22
Table 2: examples of possibilities of Topic modelling 41
Table 3: Typologies of Knowledge Maps 44
Table 4: Comparison of advantages and limits of the different information architectures 50
Table 5: Comparison of visualization tools 58
Table 6: Actions/needs matrix for enrichment and control of information by users 61
Table 7: examples of IP for knowledge creation process (Moor, et al., 2002) 65
Table 8: Criteria for system quality 67
Table 9: Possibilities of Information Filtering 76
Table 10: From the LPP to the definition of goals 79
Table 11: Book of specifications 81
Table 12: identification of items emphasized by book of specifications 84
Table 13: comparison of KM methodologies 85
Table 14: Proposed Methodology 89
Table 15: Choices of tools and methods supporting the methodology 96
Table 16: System of Performance Indicators 97
Table 17: Methodology and tools versus book of specifications 99
Table 18: Assets of Indutech's tools for the methodology 109

ACRONYMS

[redacted]

ACKNOWLEDGEMENTS

I would express my gratitude to all those who gave me the possibility to make my research internship in South Africa. I thank Alain Bernard and Nick du Preez for having established the partnership between the department IRCCYN in Nantes, and the GCC and Indutech in Stellenbosch.

I am obliged to Louis Louw, my supervisor in Indutech, and also Wilhelm Uys and Ernst Uys, two researchers of the company, whose help, stimulating suggestions and encouragement helped me in all my research works.

I would like furthermore to thank Catherine Da Cunha, Sammar Ammar-Khodja and Nicolas Perry, my professors of Centrale Nantes, who were always available to answer my question.

I thank also all my colleagues in Indutech, for their interest, their help, their kindness and all these good shared moments. Especially I was glad to meet Simon, a German intern who shares my room space, Marize and Marishka, who gave me a lot of information about South Africa, and Johanna, Heinz, Monique and Elna, who took the time to discuss with me and give me some practical information.

Finally, I would give my special thank to my family, my girlfriend and my friends, who support me during all my stay in Stellenbosch.

Stellenbosch, October 2007

Introduction

MOTIVATION

To stay competitive, companies have to innovate faster and more often. Thus, people need to interact with other people and to access information to create new knowledge. That resulted in the creation of new organizational structures, the Communities of Practice.

Due to globalization and progress in computer transmission, ways of working together are changing. Nowadays, people having the same interest or expertises wish to be connected, whatever the distance is between them and in an asynchronous environment. Hence new virtual networks have been appearing, increasing possibilities about information sharing, mass of participants… but also some difficulties for learning, communicating, developing practices and assessing the exchanged information.

In parallel, over the last years, the conversion of text into digital format has helped people to find and share knowledge more easily and more quickly. Moreover, it has improved collaboration in knowledge networks, e.g. by making it possible to identify people's expertise from the analysis of electronic publications. The abundance of information and the size of corpora (which continue to increase), however causes a lot of difficulties such as indexing a document, identifying its topic, giving its main characteristics (like its author, its research area…), and retrieving all this information.

To solve these new problems, the need for a computerized information generation (or reification) process is obvious. The fields of indexation, extraction, classification, organisation and visualization of information must be explored in order to push it to users, diffuse it and share it in a collaborative context.

I will try to answer these problems in this master thesis, in proposing a methodology to improve information generation in a community of practice. I will also think about some tools supporting this methodology.

FRAME OF MY RESEARCH WORK

For my research internship, I worked for six months in Indutech[1], a small company located in Stellenbosch, South Africa. This firm commercialises methodologies and software to help other companies to improve their knowledge management, their innovation and decision-making skills.

Indutech is developing especially three softwares:

- CAT which allows extracting and classifying concepts from corpus,
- ORGANON which provides a tool for mapping conceptual networks.
- EDEN which is an Enterprise Wide Innovation Management tool that helps project teams to manage information and knowledge along a roadmap based on project life cycle.

These tools provided a good base to understand the different processes of knowledge generation and to develop the methodology.

This work is also linked to VRL-KCiP[2], because I will make a short case study[3] on this specific network. This one is a European knowledge community in the domain of production, which groups laboratories of several different countries. It aims at improving its collaborative work, its sharing and retrieval of documents, and the identification of its members so that people can create more easily partnerships.

[1] (Indutech, 2007)
[2] (VRL-KCiP, 2007)
[3] See Appendix 1 for the scope definition of my works

THESIS PLAN

The first part introduces the background of the master thesis. Chapter I deal with communities of practice and their characteristics, whereas Chapter II insists on the evolution of CoPs towards virtual networks and the new barriers in the frame of artefact digitalization and virtualization. These observations lead to the hypothesis and the thesis problems of my subject in Chapter III.

The second part is a state of art oriented by the different highlighted problems. Chapters IV to VIII provide overview about the necessary techniques, tools and methods to develop the further methodology.

After specified more precisely goals and objectives in Chapter IX, the third part presents my propositions of a methodology to adapt knowledge generation in the context of a virtual community of practice, and the bases of a KMS to support that. I will bring some elements of justification, and explain my choices. Finally, I will propose some indicators to assess the different “actors” involved in my proposition.

In Chapter X, I will study the software from indutech and determine how they can help the implementation of the methodology. I will make some verifications in the Chapter XI, in studying theoretically the proposed methodology on a case study, and observe how it can fit the needs of a virtual network, VRL-KCiP.

Part A. Observations & Problems

Conceptual Background

Today markets have evolved towards a demand which leads and pulls supply. Accordingly organizations must face uncertainty and continuous change. They can't keep a rigid structure and a routine; they have to adapt themselves to be able to improvise solutions quickly and correctly, to respond in the better way to the needs of clients.

Moreover globalisation is another pressure which imposes modernisation of organizations. If firms began by restructuring through outsourcing and downsizing, which results in a loss of knowledge, they consider nowadays that the increased internationalisation should bring about an increased need for knowledge sharing (Kimble, et al., 2000).

To respond to this changed environment, companies create discrete groups and teams based on more fluid organizational forms such as networks and communities. The more recognized structures are called Communities of Practice (CoP). These ones, because of the globalization and outsourcing, tend to become virtual.

This first section will define Community of Practice and its knowledge background in chapter I. We will focus on the knowledge life cycle in CoP and we will attempt to characterize them with two models, the Legitimate Peripheral Participation and the duality participation/reification.

Then we will present, in Chapter II, the new needs of these networks and their transformation in a global frame, based on two main evolutions, artefact digitalization and virtualization. If these progresses provide some advantages, they also cause the appearance of some limits and barriers for the functioning of CoPs.

At long last, we will give the hypothesis of our study in the third chapter, and introduce the thesis problem in organizing the questions of overcoming the barriers emphasized in Chapter II around the models presented in Chapter I.

CHAPTER I. COMMUNITIES OF PRACTICE & KNOWLEDGE BACKGROUND

Keywords: Knowledge, Artefacts, Life cycle, Communities of Practice, Legitimate Peripheral Participation, Reification

Before presenting Communities of Practice and their characteristics, we will discuss some preliminary notions about knowledge and its life cycle. That will allow further to emphasize the impacts of CoP on this knowledge life cycle.

I.1. Preliminary background: towards a knowledge life cycle

For centuries philosophers and academics have debated the meaning and role of knowledge. Yet, knowledge has proven to be an evasive term. The inability of researchers to unequivocally define knowledge illustrates this point" (Croassdell, et al., 2003). Thus we will attempt to clarify this concept and show its typologies and the fundamental role of artefacts. Then we will discuss about the knowledge life cycle.

I.1.1. Between epistemology and systemic approach

The literature emphasizes two main approaches. On one had, epistemology claims the importance of human interpretation in knowledge creation as well as the technique as a means to externalize memory (Bachimont, 1996; Charlet, 2002). On the other hand systemic science assumes knowledge is at the top of a pyramidal triptych data-information-knowledge, and characterizes it by its temporality and a context of use (Ermine, 1996; Poitou, 1996; Candlot, 2006)

According both these two approaches, we can acknowledge some principles about knowledge: It is based on information, it is human, it is temporary, it needs an interpretative endeavour in a context, and it creates action.

Thus, we propose the following definition:

Knowledge is a temporary comprehension, resulting from human interpretation of information in a specific context and a constructive process of modelling, and is adapted for transformation into action.

I.1.2. Typologies of knowledge and artefacts

KM literature has several different typologies, but acknowledges two widely used types of knowledge: tacit and explicit knowledge. This classification deals with the problem of representation and formalization, in order to use it easily and act with it.

- Tacit Knowledge

(Polanyi, 1966; Canard, et al., 2004) explains that tacit knowledge is personal, intuitive and not articulated. It is hard to formalise and transmit it to others. So this knowledge is hardly to codify, and to be transmitted (Sekkat, et al., 2005).

- Explicit Knowledge

(Nonaka, et al., 2000) argue also that *"explicit knowledge can be expressed in formal and systematic language and shared in the form of data, scientific formulae specifications, manuals and such like…"* Consequently, we could express explicit knowledge with codes and symbols, which permit to communicate and transmit it.

Thus, one of the most important challenges in knowledge science is to transform tacit in explicit knowledge, especially to keep a memory of it and raise its common understanding.

- The concept of artefact

(Jaime, 2005) defines artefact as *"an element having a material form (or a virtual form, as it can exist only in a computer system) which can convey a part of knowledge held by its author, provided that its receiver knows the context in which it was conceived and has the necessary knowledge for its interpretation".*

Thus, artefacts (and so documents) enable to solve a part of problem of representation of tacit knowledge, relatively to explicit knowledge. It is a tangible representation of tacit knowledge, and it can be a way to transform that one into explicit knowledge, under a codifiable form. We can also consider, according the epistemological approach, that it is the technical tool to externalize knowledge.

I.1.3. Knowledge life cycle

(Judelman, 2004) analyses the processes between data, information and knowledge in the following table. He emphasizes especially the collection and the analysis of data, which are then transformed into information thanks organization and representation.

	data	information	knowledge
processes (along continuum)	collect analyze	organize represent	contextualize apply
properties	quantifiable disordered raw	processed structured patterned	understood verified discussed
examples	spreadsheets timetables facts	visualizations reports instructions	actions processes stories

Figure 1: Processes in systemic view (Judelman, 2004)

(Zimmermann, et al., 2002) proposed also a model based on the systemic hierarchy of data-information-knowledge, in which relations appear. These one are represented on the following figure:

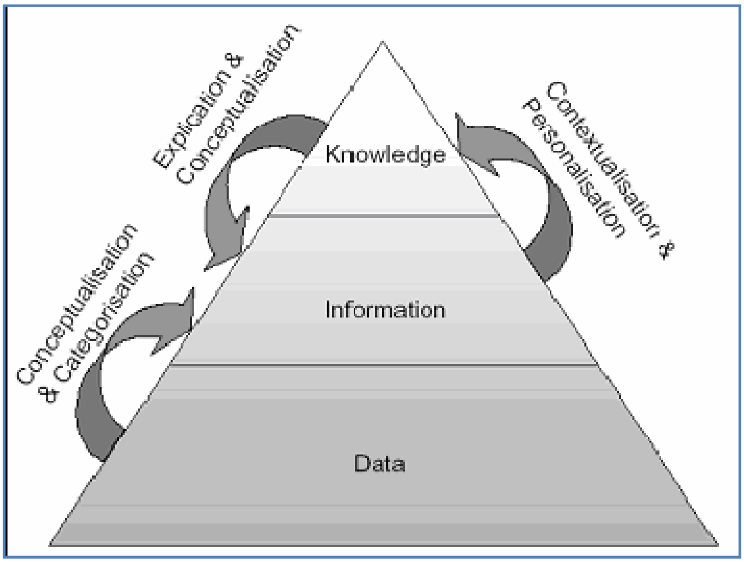


Figure 2: From data to knowledge and from knowledge to information (Zimmermann, et al., 2002)

Not only there is an ascendant process, from data to knowledge (with conceptualisation and contextualisation activities), but we point out a descendant link, from knowledge to information (with explanation).

In other respects, (Nonaka, et al., 2000) bring forward another model, the SECI principle, which presents a knowledge life based on the transformation of tacit and explicit knowledge.

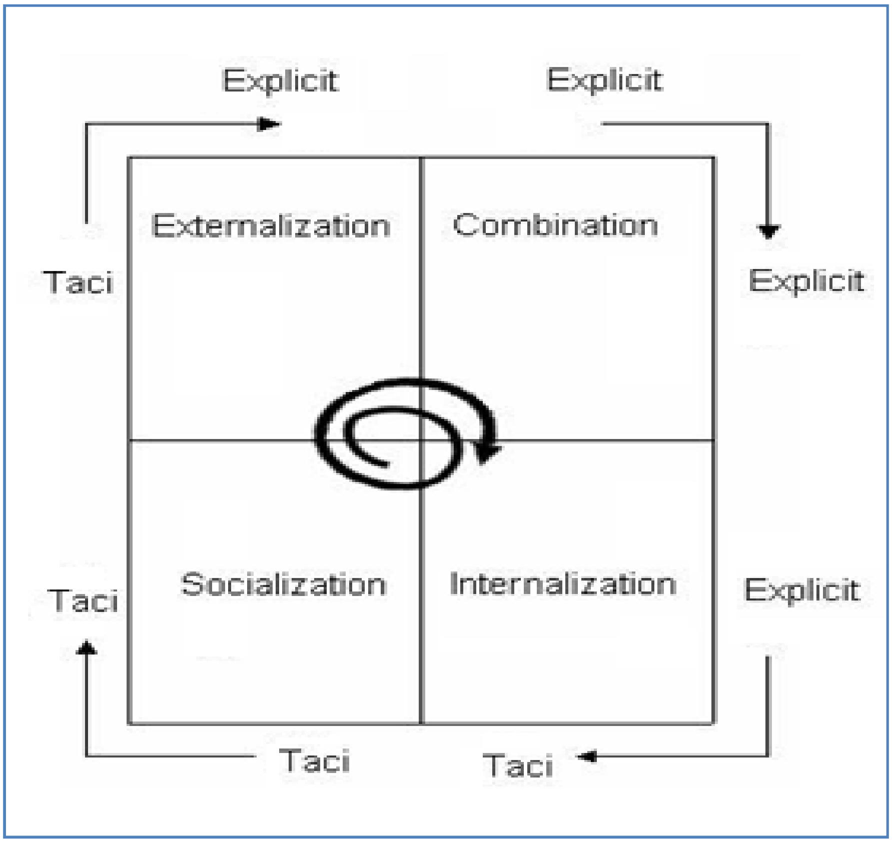


Figure 3: SECI Model (Nonaka, et al., 2000)

- Externalization is the process of converting tacit knowledge into explicit knowledge (called *"conceptual knowledge"*), such as metaphors, analogies, concepts, hypotheses
- Combination articulates explicit knowledge into more complex and systematic sets of explicit knowledge, called *"systemic knowledge"*. Examples of such a conversion process are sorting, adding, combining, modelling and categorizing explicit knowledge
- Internalization is the process of turning explicit knowledge into tacit knowledge. Internalization produces *"operational knowledge",* for example by training or learning.
- Socialization is the process of creating new tacit knowledge, such as shared mental models and skills, out of existing tacit knowledge through shared experiences, for example in informal social meetings. The resulting tacit knowledge is also called *"sympathized knowledge"*.

Model1

From these different points of view, we are going now to try to mix them. We use as draft a scheme of the SECI process drawn by (Huang, et al., 2004)

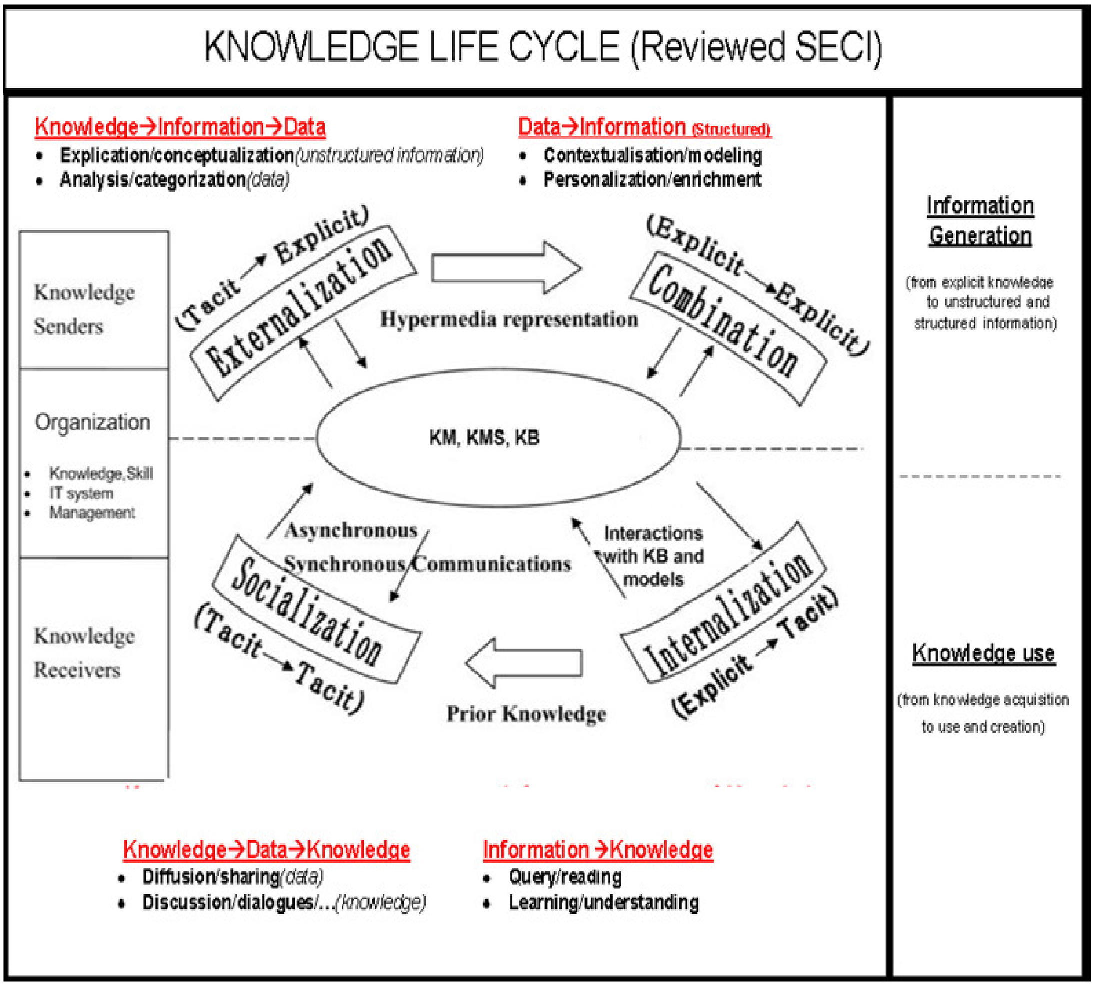


Figure 4: proposition of a divided knowledge life cycle combining SECI view and systemic view

As we can see, SECI is a robust model taking account of all the actions during knowledge life cycle, and it could be divided in two flows:

An information generation process, where:

1. Knowledge is explicated by the use of artefacts, which transform it into the form of unstructured information (knowledge → information)
2. This first step of knowledge externalization can be continued with a categorization of unstructured information in structured data, for instance in analysing topics of documents and putting them in some data tables (information → data)
3. This data is combined and enriched with semantic and logic so as to give modelled information, through information structure and representation (data→information)

A process of transformation and use, where:

4. Knowledge is acquired from unstructured and structured information by acting on the explicit sources (information→knowledge),
5. Then it is diffused in order to teach or learn (knowledge → data)
6. Finally it is discussed by people to create new knowledge (data→knowledge).

Thus, knowledge is push towards knowledge users, then it used by them to innovate and to collaborate.

The systemic approach is focused more on a low level, inside the SECI life cycle.

I.2. Communities of Practice: conceptual presentation

I.2.1. Definition

The terms Communities of Practice was coined by (Lave, et al., 1991) to describe an activity system that includes individuals, who are united in action and in the meaning that action has for them and for the larger collective. CoPs are not formal structures such as departments or project teams.

We keep a general definition of Community of Practice (CoP) proposed by (Wenger, et al., 2002):

CoPs are *"groups of people who share a concern, a set of problems, or passion about a topic, and who deepen their knowledge and expertise in this area by interacting on an ongoing basis".*

I.2.2. Typologies

(Archer, 2006) classifies CoPs in four types (sorted in two groups)[4]:

Communities of practice
a. entirely within individual organizations,
b. spanning organizations that are linked through mergers, acquisitions, or by formal business partnerships (network organizations[5]),

Networks of practice
c. formal networks that span organizations but are not part of other formal relationships,
d. self-organizing networks of individuals with ad hoc relationships and no formal ties.

[4] Details of this classification will be found in appendix 2
[5] for example, a supply chain is a network organization

The main difference between Communities of Practice and Networks of Practice is the fact that these last ones are stand-alone and not adjunct to more general agreements and contracts. The activities of Networks of Practice are usually covered by blanket agreements that assign intellectual property rights to the network and not to members of individual organizations.

Thus CoPs can take several forms, which depend on the relationships between organizations where members come from. The different types of communities are due to the contradiction of knowledge capitalization: creating new knowledge in sharing it, but also protecting strategic knowledge in multi-organizational frame. This is an important question before initiating a CoP.

I.2.3. Roles of CoPs

These networks are established in order to build strategic capabilities within the organization by leveraging learning and knowledge sharing (Prusak, et al., 1999). We will now study what they bring to individuals and organizations.

Individuals as community members profit directly from their participation in the community.

- by communicating frequently, the community members develop a common language and a collective knowledge base.
- the personal knowledge of the community members is increased, and new competences are gained which allows to improve performance (Wenger, et al., 2002).
- Due to advanced competences, community members are regarded as experts in a specific field which in turn leads to a higher reputation within the organization. This has a positive impact on their work satisfaction (Schoen, 2001).

Strategic advantages for an organization result, above all, from community impacts on the organizational level, with the externalization of knowledge.

- A common knowledge base is created also at the organizational level. Existing know-how is improved, and new organizational competences are developed (Tsai, et al., 1998).
- Communities of practice enhance the creative capacity and, by this, the innovative capability of the organization (Brown, et al., 1991).
- Resource savings result because CoPs may also decrease training periods for new employees as well as help to avoid double work.

- Optimized and accelerated processes together with the developed knowledge base will potentially lead to higher customer satisfaction, as customer needs can be addressed in a more flexible manner (Lesser, et al., 2001).
- Communities of practice can change the existing organizational culture in a favourable way (with the apparition of a common language and a positive and willing attitude towards knowledge sharing).

I.2.4. Needs of CoPs

We have seen knowledge sharing and communication are prominent in these structures. So as to innovate, it is obvious that two main needs in CoP are:

- the retrieval of information (to learn, to acquire, to internalize knowledge)
- and the retrieval of people(to collaborate, to communicate),

I.3. Two Wenger's models and knowledge life cycle in CoPs

Over the time, Wenger has proposed two models for the description of CoPs and the interaction inside them. We are going to present these models in order to characterize CoP's functioning and show how CoP interact with knowledge life cycle.

II.3.1. Legitimate Peripheral Participation

As (Lave, et al., 1991) emphasize, the acquisition of knowledge in CoP is a social process where people can participate in communal learning at different levels depending on their level of authority or seniority. Moreover it is situated: learning must be done in a contextualized environment.

So this situated learning process, central to the notion of CoP and by which a newcomer learns from the group, either by communicating either by interacting with media, is called Legitimate Peripheral Participation.

LPP can define evolution of a member in a CoP. *"A new member of the community moves from periphery to full participation in the community"* (Hildreth, et al., 2006). Initially its activities and may be restricted to simply gathering domain knowledge, in reading or talking with *"experts"*. Later the newcomer may become involved with gaining knowledge associated with the specific work practices of the community, for example in doing basic tasks. Gradually, as the newcomer learns, the tasks will become more complicated and the newcomer becomes an old-timer and is recognised as a source of authority by its members.

Two status for CoP's members

From this analysis, we can identify two statuses for CoP's members:

- user can be ***passive***, as an ***observer***, i.e. he just gathers information and discuss with people to learn
- user can be ***active***, as a ***contributor***, and bring some information, create knowledge, when he has acquired the status of expert.

LPP Interpretation

Lave and Wenger's view: LPP is both complex and composite, and must not be divided to understand all its aspects, according to (Lave, et al., 1991). It shows indeed the dynamic linked flows for gradually learning, for becoming recognized expert, for increasing his possibilities of participation…

Hildreth and Kimble's view: However, to explain it and decrease its complexity, (Hildreth, et al., 2006) tried to separate the concept in three aspects: legitimacy, periphery and participation.

- Legitimacy refers to the power and the authority relations in the community.
- Periphery refers to the individual's social rather than physical periphery in relation to the community.
- This in turn is dependent on their history of participation in the group and the expectation of their future participation in and interaction with the community.

Personal view: Between these two approaches, composite for (Lave, et al., 1991) and granular for (Hildreth, et al., 2006), we propose another analysis of the concept.

If we attempt to divide LPP by pairs, this emphasizes three relevant points:

- peripheral participation seems to refer to:
    - progressive learning (CoP's member is an observer), when users acquire knowledge about a domain and about other members, in gathering information and in discussing with the others,
    - progressive interactions (CoP's member is a contributor), when people can gradually interact.
- legitimate participation refers to authorized interactions that a user makes with the others or the resources information.
- peripheral legitimacy determines the progressive degree of expertise of participating users, which authorize interaction.

Nevertheless, if progressive learning is obviously linked to peripheral participation (because the fact to participate from periphery to expertise is clearly a gradual learning), the notion of progressive interactions is not completely defined by peripheral participation. Indeed, to participate more and become a contributor, we have seen with (Hildreth, et al., 2006) that user needs to become expert, i.e. to be legitimate.

This loose sense could be solved. By transitivity and combination, legitimate participation and peripheral legitimacy underlie that peripheral participation refers to progressive interactions. Indeed, the legitimate participation uses the peripheral legitimacy to authorize gradually the activities of CoP's members.

Model2

So we will assume with our interpretation by pairs that:

- ***Peripheral Participation*** refers to ***progressive learning***,
- ***Legitimate Participation*** refers to ***authorized interactions***.
- ***Peripheral Legitimacy*** refers to the ***progressive degree of expertise*** of participants.
- ***Legitimate Participation*** and ***Legitimate Periphery*** underlie ***progressive interactions***, the gradual activities being determined by peripheral legitimacy.

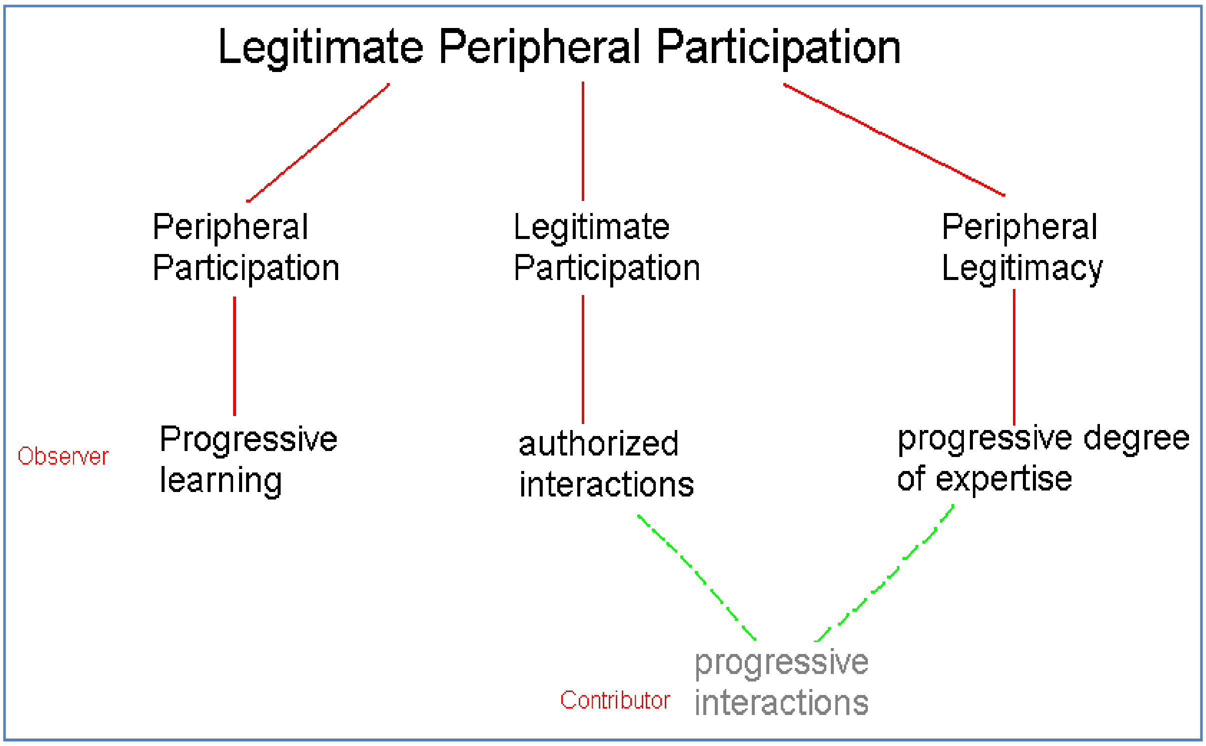


Figure 5: Division by pairs of LPP model

II.3.2. Participation-reification duality and modified knowledge life cycle

More recently, (Wenger, 1998) reviewed their model, and pointed out other aspects of CoP learning in using inherent tension between dual notions. The more relevant point seems to be the participation and reification duality.

Reification

Reification defines the process to capture knowledge in a physical form, under a document, a model, a scheme.

According to (Young, 1992), it is the "treatment of an analytic or abstract relationship as though it were a concrete entity".

(Krippendorff, 1986) contends that it is "the process of regarding something abstract as a material entity, Whitehead's "fallacy of misplaced concreteness," e.g., the mistake of confusing a system, which is a construct, with the physical entity described in its terms".

Actually, we can distinguish some different meanings of reification:

- In ***knowledge representation***, reification is used to represent facts that must then be manipulated in some way.
- In ***computer sciences***, it makes a data model for a previously abstract concept.
- In ***linguistics***, it transforms statement, actions and events in quantifiable variables.

Thus, reification is for instance when people transform their knowledge into explicit information with artefact, and when a map or a scheme, drawn manually by people or automatically by computer, represents the information held in a corpus, etc.

Participation

Participation focuses more on the knowledge acquisition and the mutuality between users.

Indeed, the participation represents more the users' actions and interactions between users, especially the task of gathering information from documents or people (learning) then the discussions when new knowledge is created among CoP members (exchanging and innovating).

This duality takes place in a context taking account of experience of users and organizations, world environment (the constraints and the limits which rule CoPs), and a negotiated meaning (i.e. common understanding and acceptation).

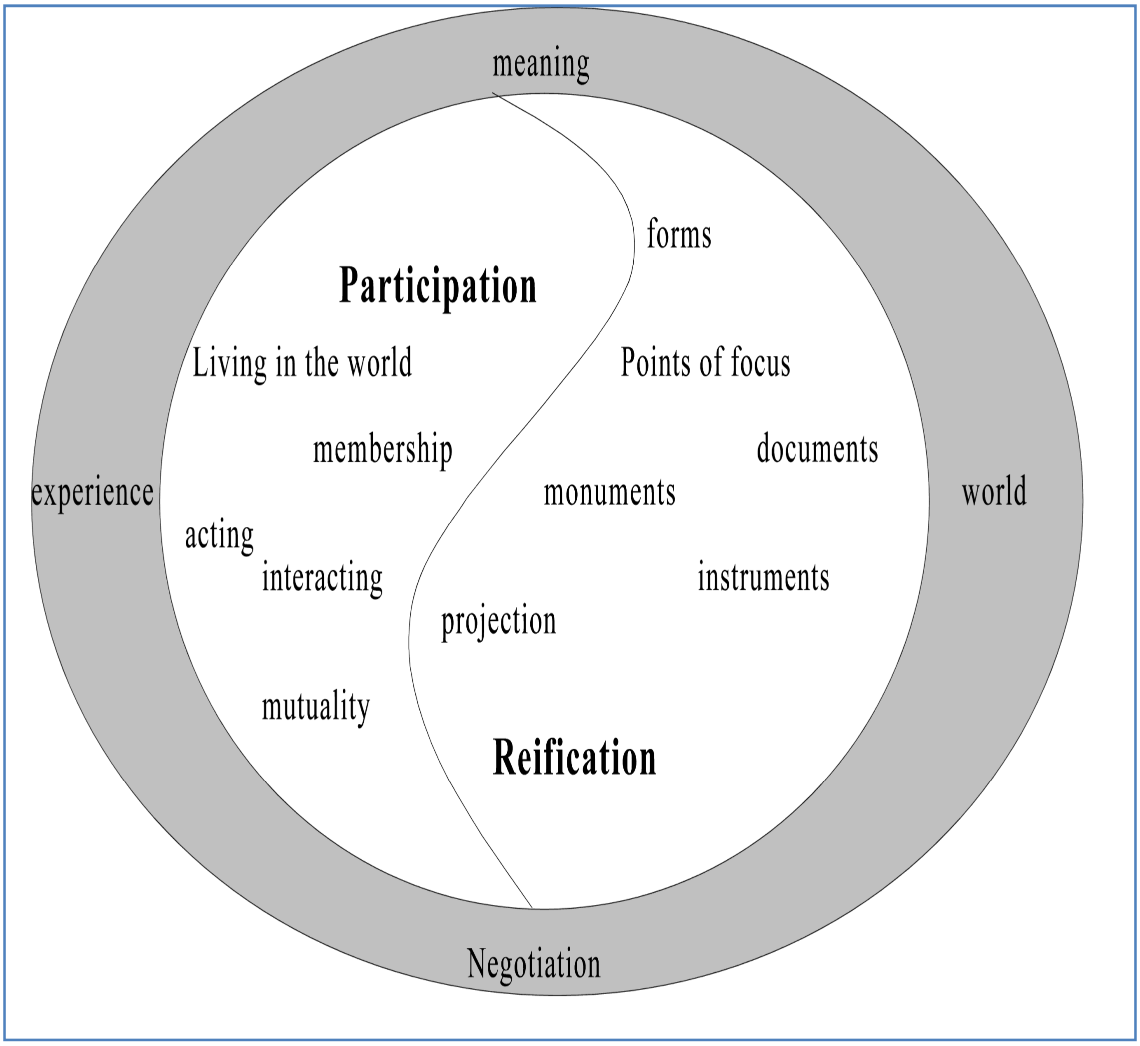


Figure 6: Participation/reification duality (Wenger, 1998)

This duality emphasizes the significant role of media (through reification), and the importance of actions on media and communication (participation).

Furthermore, this model reminds strangely the decomposition of SECI in two flows proposed in Fig.4. In addition it introduces interactions between them.

Model3

So new labels on reviewed SECI model may be given:

- Reification refers to the information generation process: people can transfer their knowledge into artefacts (document), map or scheme can represent knowledge of a corpus, etc.
- Participation refers to the process of acquisition and use,
- These two flows are dual and porous in the knowledge life cycle.

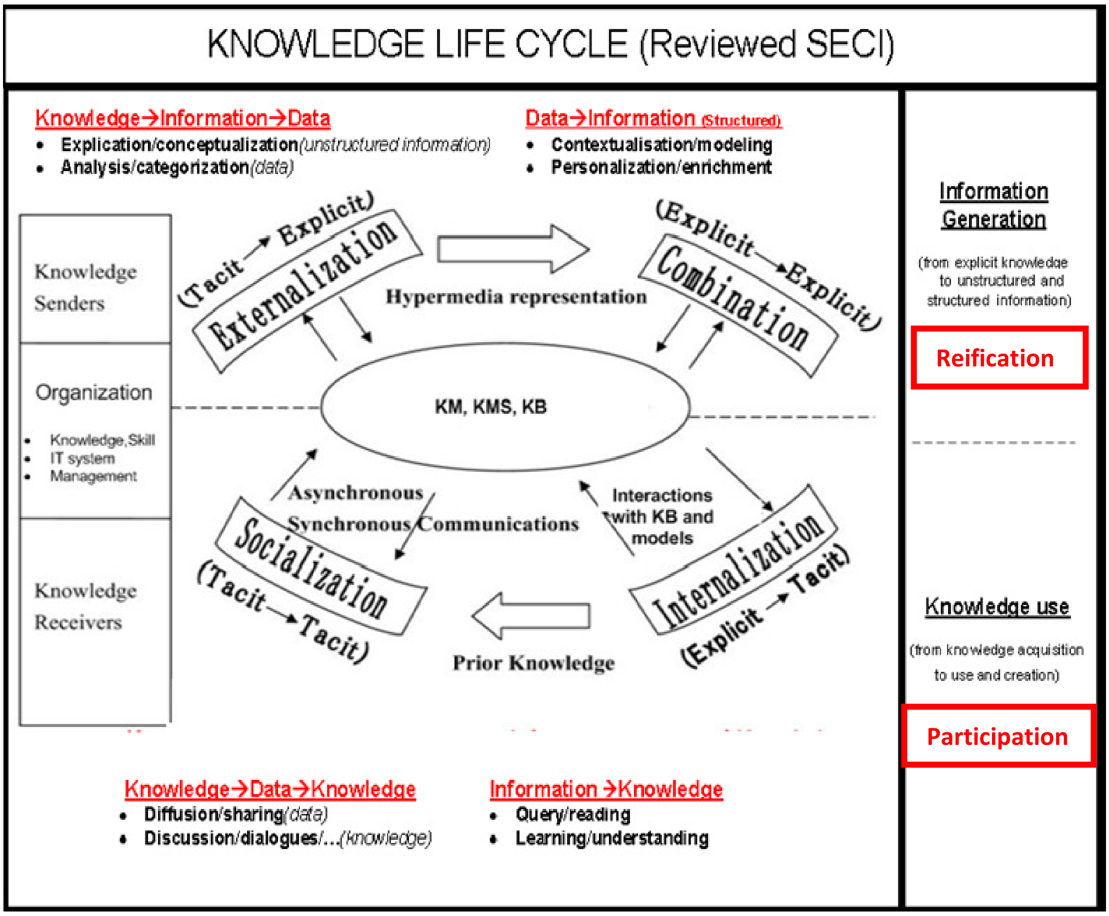


Figure 7: Divided SECI model with the Wenger's Duality

Moreover, this model shows obviously that retrieving information about knowledge domains and people who are expert in these domains are tremendous needs for the participation process, in order to learn (acquire) and collaborate (discuss) for innovation.

To respond to these needs, the step of reification is eminent, in pushing this relevant information to CoP's members.

I.3. Conclusion

We have studied in this chapter the structure, the characteristics and the principle of CoPs, as well as the knowledge life cycle, on which CoPs interact.

CoPs are the new organizational vectors for sharing and creating knowledge. Overcoming rigid and hierarchy organizations, they propose informal structure to group people around a topic of common interest, in order to innovate and solve problems.

They can be characterized by:

- the tremendous needs for retrieval of relevant information and expert people, in order to learn, collaborate and create new knowledge.
- Two models of Wenger: the Legitimate Peripheral Participation and the duality participation/reification.
    - The LPP shows the gradual process of CoP participants and emphasizes the significant questions about its legitimacy and its participation.
    - The duality accounts better for the influence of CoP in the knowledge life cycle. Thus the benchmark SECI model of Nonaka can be explained and divided in a reification flow and a participation flow (according Wenger duality), where knowledge is transformed many times into information or data forms (according systemic view).

We have defined and characterized CoP and its functioning. We will now use these characteristics so as to approach its evolution and study the potential barriers in the frame of technological progresses, especially artefact digitalization and CoP virtualization.

CHAPTER II. EVOLUTION OF COPS, BARRIERS & LIMITS

Keywords: CoP's evolution, virtualization, digitalization, barriers

We are going now to study CoPs in their current frame, i.e. in a global, virtual and digital environment. We will use our previous remarks about the definitions and the principles of CoPs to point out and analyse the limits and the barriers in with new background.

II.1. CoPs, technological progresses and questions

II.1.1. Virtualisation and growing importance of artefacts

Because of business globalization, many organizations now work in a geographically and temporally distributed international environment. Thus CoPs tends to become virtual in an adaptive endeavour. Wenger assumes for this mutation that virtual CoPs must evolve towards a constellation of interrelated CoPs (Wenger, 1998), where fast knowledge diffusion and assimilation of social networks, and creation of new knowledge and meaning of CoPs, would be combined.

Q1

But this combination looks complex, and raises the question: can CoPs continue to operate in such an environment? Can a CoP be virtual? (Kimble, et al., 2000)

Difficulties for the formation such a virtual CoP are the development of a common language, generally coming from interactions between members, like face-to-face exchanges. (Brown, et al., 2000) give a first trend in showing that documents – from newspaper to mailing lists – can generate a common language over these wider networks. Furthermore, the interactions between user and documents enable to characterize CoP users.

Thus in order to generate the common background necessary for CoP, the role of document is pregnant.

This point of view emphasizes the idea that artefact, and so the process of reification in knowledge life cycle, become prominent in virtual CoPs.

II.1.2. artefact digitalization

Artefact has taken many forms over times. (Judelman, 2004) describes this evolution in the figure below.

newest communication technology (with beginning of relevant time period)	cultural characteristics	language characteristics
speech (100,000 BC)	complex institutions religion technique symbols music hunting and gathering learning by imitation and experience	oral transmission myths rituals memory inscribed in physical matter (sculptures, tools)
ideograms (3200 BC)	cities state law complex religions with a clergy agriculture cattle breeding schools	visual symbolic systems representing concepts
alphabet (700 BC)	citizenship monotheisms buddhism money commerce universities science	universal writing system representing sounds
mass media (1500) *printing press* *photography* *telegraphy* *telephony* *cinema* *radio* *television*	modern democracies reform liberalism human rights industrial revolution world market capitalism libraries experimental natural sciences	technical self-reproduction of images and sounds
digital media (cyberspace) (1970) *media convergence* *interactivity* *global network*	megalopolis network planetary cyberdemocracy religious convergence biosphere's evolution monitoring knowledge economy nature of information	ubiquity of signs interconnection of messages autonomous activity of software

Table 1: evolution of knowledge transmission and learning (Judelman, 2004)

Digitalization is obviously the more recent progress. After mass media, which supported group communication, digital media has provided a means to adapt human exchanges in a global framework.

Many possibilities have appeared:

- possibilities of interactions on documents, like creation, editing or modification, ease give a democratic way to make artefacts;
- facilities for access and transmission help users in diffusion and knowledge sharing;
- progresses in compression and increasing size capacity have raised document storage.

However it opened also a Pandora box: the abundance of information brings about issues for the generation process, especially for indexing and categorizing unstructured corpora.

Moreover, we have explained in the previous part that documents are playing a growing role in new virtual CoPs in order to generate a common background for learning and collaboration.

Q2 **Is it possible, with the abundance of digital documents, to have an accurate common language for CoPs?**

II.1.3. Social networks, blogs and wikis

As the debate about the nature of « virtual » CoPs got underway, the rapid diffusion of Internet-based networking technologies was accelerating the development of new forms of community: the social networks (Hildreth, et al., 2006). These one are virtual networks of persons linked by social relationships (like hobbies, work activities, family…). They have a global framework and are supported by unsynchronized platforms, like blogs or wikis, two tools from Web2.0.

Thus social networks are a fine example to analyse the impacts of virtual environment on communities, and so on virtual CoPs.

According to (O'Reilly, 2005) *"Web 2.0 is the network as platform, spanning all connected devices; Web 2.0 applications make the advantages of that platform: consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an 'architecture of participation'…*

A wiki is a website enabling users to not only create and add content, but also edit content. It provides a more effective way of information exchange through collaborative effort. *"A defining characteristic of wiki technology is the ease with which pages can be created and updated"* (GoodwinJones, 2003). A Knowledge Management system would benefit from wiki technology due to its ease of use, its ability to capture knowledge in a shared and growing repository, its wide accessibility options (via a web browser) and its *"Architecture of participation".* Wikis also provide for flexibility in decentralized organizations by giving the knowledge workers the tools to react quickly to changing situations (GoodwinJones, 2003).

Blogs are *"the collaborative environment which has sparked the most intense interest in recent years"* (GoodwinJones, 2003). Most blogs are more than just online journals; they are interactive while being structured knowledge repositories. Blogs offers a set of tools for users to post comments and share their knowledge with other readers. And because of the structured nature of blogs, the knowledge that is captured is easily found and remixed by other users.

Q3

The ease of creation and participation are emphasized in social networks and the Web2.0 platforms. However, how strong is the socialization process of SECI model, i.e. the interactions between users in communal learning? Does the participation in virtual communities support LPP the acquisition of common practices?

II.2. Barriers and limits of virtual and digital CoPs

II.2.1. Barriers: new CoPs and LPP Process

- Practices and communication are more limited in these new environments.
    - the knowledge about the others decreases, on account of the abundance of members
    - face-to-face exchanges and story-telling become very difficult to manage in global and unsynchronised frame (Hildreth, et al., 2006).
    - mass of information can lose people in learning activities, and avoid them an easy access to documents and people.

The progressive learning and so the peripheral participation becomes infringed with virtualization and digitalization.

- Moreover, the need for identifying and qualifying members is very important in new virtual CoPs, where people have no time to learn about other users.
    - It is harder to know if a user if reliable or not
    - Another difficulty is to reckon who may participate and how

The legitimacy of periphery and participation are at stake and are interfered with the virtualization.

- Finally, the context is essential or learning, and according (Lave, et al., 1991), LPP is situated.

  So there is a risk to lose the context in making CoP virtual.

Barriers

We can resume all these barriers by the lacks in LPP process in order to create confident relationships, participation and motivation in CoPs, where people can't discuss together and where actions are done in a global and unsynchronized framework.

II.2.2. Limits: socialization versus internalization in learning process

Social networks seem to provide a good example of solution for CoP virtualisation. Indeed the rapid diffusion of Internet-based networking technologies was accelerating the development of new forms of community. However, according (Hildreth, et al., 2006), it has also made increasingly difficult for people to know the scope and range of their *"virtual"* social networks.

Thus, the acknowledgement of the legitimacy of other members in a virtual environment is very hard, despite the apparition of social networks.

Moreover, in looking at the social communities organized around web2.0 tools like blog or wiki, we can observe that mass of information and abundance of members limit collaborative exchange and social learning. Indeed, according to Jimmy Wales, the creator of Wikipedia, *"the most active 2%, which is 1400 people, have done 73.4% of all the edits."The remaining 25% of edits", were from "people who [are] contributing a minor change"*. That tends to show the *"1% rules"* described by (Mons, 2006). Likewise, a recent survey shows only 13% of internet users are creators and 19% interact with these creations (Li, 2007).

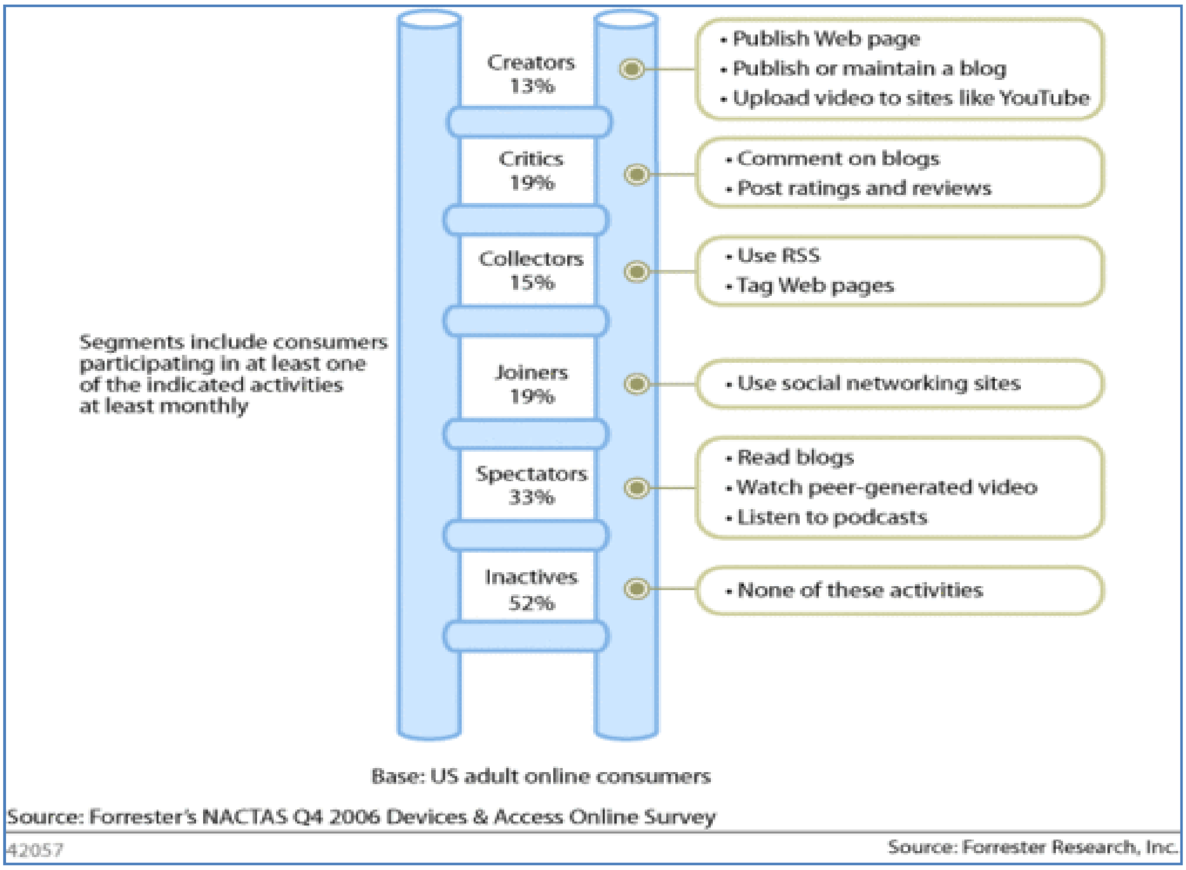


Figure 8: participation in virtual environment (Li, 2007)

These figures outline the passivity of users. People pick information, but few of them are active, responsible for the creation of content or interacting with creations.

Thus, the socialization process in the participation flow presented in Fig.6 looks weak for learning in virtual environment. Learning seems to occur more in the internalization process (in acting with the knowledge bases) than during social interactions (discussions…).

Limits

Thus, the socialization process in the participation flow presented in Fig.6 looks weak for learning in virtual environment. Learning seems to occur more in the internalization process (in acting with the knowledge bases) than during social interactions (discussions…).

Because people are more passive, it is important to push information towards them, and so to improve the reification process.

II.3. Conclusion

With the changes of globalization and the need for flexibility, CoPs must have been developed towards unsynchronized and virtual environment. If this virtual aspect facilitates transmission of knowledge and increases possibilities (e.g. the number of members), it brings also some difficulties, especially concerning the LPP process, i.e. the acquisition of common practices, the learning and the legitimacy of CoP's members.

New virtual CoPs are facing to a dilemma: adapting them to globalization in boosting the diffusion of knowledge and in keeping the process of creation of common practices. Documents bring some answers, in providing a means to create a common background, showing the significance of the reification process.

However, virtualization and digitalization cause some huge problems and underscore the dilemma: the LPP process presents some barriers in virtual CoPs, whereas the collaborative learning decreases, emphasizing the limits of socialization in virtual environment and a new need for pushing information towards users.

With the aid of the previous observations, we have emphasized the limits and the barriers of CoPs in a new global, virtual and digital background. These limits should be obviously overcome. We are going now to formalize the problem in the following chapter.

CHAPTER III. HYPOTHESIS & THESIS PROBLEMS

III.1. Hypothesis, framework of thesis problems

CoPs evolve and must face to a new context, a global environment and two main technological evolutions: virtualization and digitalization.

Virtualization, supporting geographically distributed environment and unsynchronized communication, shows the difficulties to keep the own essence of CoPs: the sharing of practices and the development of a common language, becoming complex with the nature of communication (decrease of face-to-face exchanges, not situated…)

Documents seem to balance this lack in creating the necessary common background, but they present also some issues due to the abundance of information which noises the practices and the shared language.

From these evolutions, barriers and limits appear in the functioning of new virtual CoPs, we need to overcome. Emphasized by the Model 2, we have observed that they concern the behaviour of the LPP process introduced by Wenger and Lave.

Hypo1

Hypothesis1: the LPP model allows for analysing the barriers and limits in new virtual CoP, and could be used to overcome them, in approaching the problem according the three points emphasized

On the other hand, social learning decreases inside CoPs, due to the structural constraint of globalisation. The acquisition of shared and common practices is more concentrated in the internalization process. According Wenger's duality and the limits, reification should then be increased, in order to respect the inherent tension between participation and reification. In addition, following the model 3 of the reviewed SECI model, the improvement of reification would ease the internalization and the development of a common language.

Hypo2

Hypothesis2: reification improvement could reduce the barriers concerning the LPP process and the development of common practices.

III.2. Thesis problems

Consequently to the hypothesis, we will focus our problem on the barriers emphasized by the LPP process in virtual CoPs, and we will limit our study on reification.

PB

General problem: How can we apply reification process in a virtual Community of Practice, by interacting with documents knowledge base and in respecting the Legitimate Peripheral Participation?

Following the second hypothesis, this main issue could be decomposed according the LPP interpretation presented in model 3.

Problem of information mass and decrease of participation suggests the need for pushing information towards CoP's members, in order to ease learning about domain and other members. Thus, one needs to automate reification to boost the peripheral participation, i.e. the progressive learning.

Pb1

Problem1: How can we automate reification to ease Communities of Practice's progressive learning about information and people?

Because knowledge is dynamic and interactive, this automation cannot be sufficient and participation must be organized in order to control, add or modify some information, when you are legitimate. Interactive enrichment of automated reification is necessary.

Pb2

Problem 2: How can we enrich and control documents knowledge base and reification process, when legitimate?

The expertise degree of user changes continuously. So enrichment must be constantly monitored according the peripheral legitimacy of CoP's members, in order to authorize them to participate.

Pb3

Problem 3: How can we assess and authorize users and their actions during the reification process?

III.3. Conclusion

These thesis problems are based especially on the combination of the knowledge life cycle (reviewed SECI) and two models of Wenger about the functioning of Communities of Practice, the Legitimate Peripheral Participation and the duality reification/participation.

This approach enabled us to understand better the evolution of CoPs into a global, unsynchronized, virtual and digital environment. Thanks to them we have also highlighted the barriers and the limits of these new virtual CoPs, emphasizing that the issues interfere rather with the LPP process and that the endeavour must be done on the reification.

It is primordial to keep in mind the limited scope of the study, which focuses solely on the flow of reification, and excludes the collaborative work for creating new knowledge and innovating.

Thus, these problems seem to lead towards a methodology and the choice of some adapted tools to apply an assessed interactive automated reification in virtual CoPs.

To conclude this part, we will propose a scheme summarizing our problem.

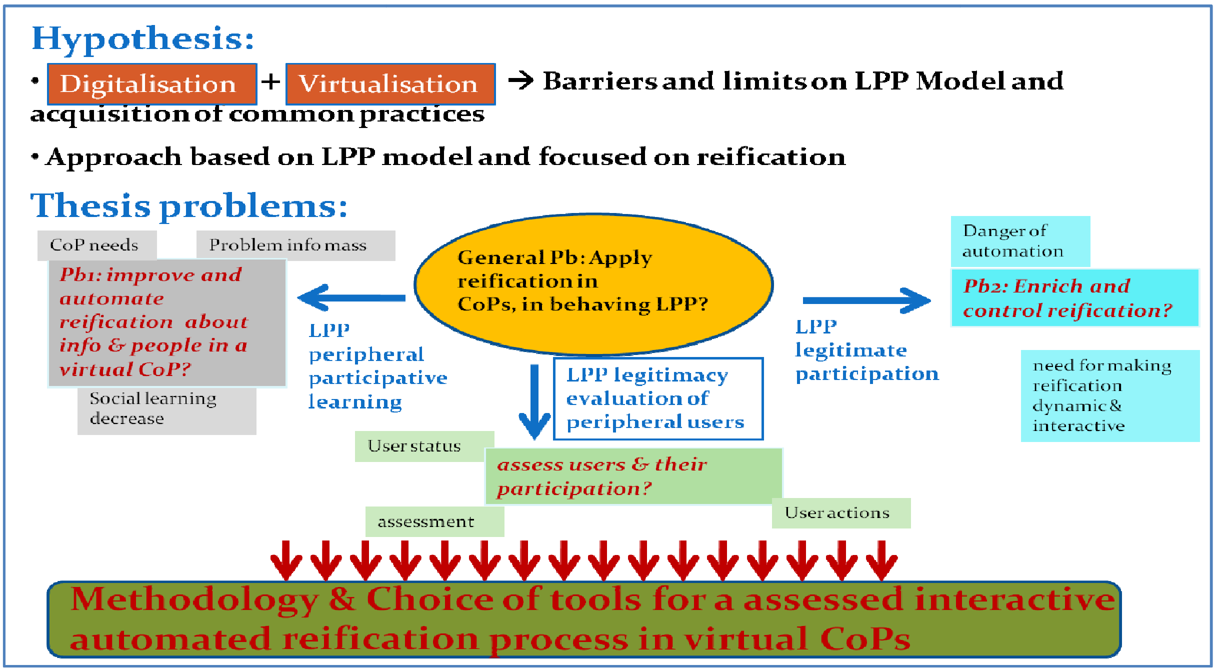


Figure 9: Summary of thesis problem

We are going now to make a state of art of the different techniques and methods providing a solution for the differents sub-problems presented above. This “toolbox” will then be used in order to build our methodology and justify our choices for the tools supporting it.

Part B. State of Art

Techniques, tools and methods

This part in our thesis will give a toolbox about reification, participation and assessment, so as to solve our problems presented previously.

The first section deals with the automation of reification enounced in Problem1: we will explain the processes and the methods for extraction, organization and visualization, which push information towards user and ease cognition and future acquisition.

The second section will be related to Problem2 and Problem3, focusing on enrichment, participation, control and assessment of automation, and it will show that these notions are narrowly linked.

This state of art will be necessary for proposing a methodology and suggesting some tools supporting it in order to solve the main problem of our thesis (PB).

Section 1: Computerized reification

As stated by the logician Alfred North Whitehead:

> *Human knowledge is a process of approximation. In the focus of experience, there is comparative clarity. But the discrimination of this clarity leads into the penumbral background. There are always questions left over. The problem is to discriminate exactly what we know vaguely.*

Understanding and representing all the things (in externalizing and combining them) of our environment is a perpetual human quest. Yet, in antic and medieval times, people try to classify and model the human being, the nature…

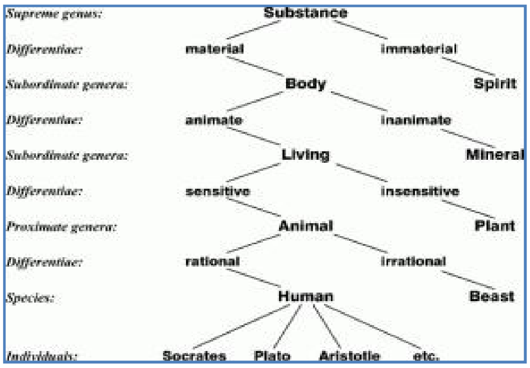


Figure 10: Tree of Porphyry drawn by Peter of Spain (1329)

This will to explain and conceptualize the world joins the concept of reification. This process was shown to be divided in two parts, externalization and combination. Obviously we can not automate the understanding and transformation of explicit knowledge (help by people) into artefacts (information form), which must be done by a human process. However, this externalization of knowledge can be completed and enriched by automation.

Indeed, the new breakthroughs of topic extraction in increasing-size corpora, and representation of extracted information provide larger possibilities. Whereas computer science brought its power and its logic to help people for artefacts analysis, philosophical ontology became new computer systems and figures of Peter of Spain were replaced by knowledge visualization software.

Thus, we will attempt in this state of art to give the tools allowing this computerized reification:

- analysis, summarization or conceptual categorization of corpora may be automated by a computerized way, with content analysis which transforms unstructured information in structured data,
- this structured data can be then organized in including logic and semantic, with information structure, like ontology or topic maps.
- finally, structure can be visualized, to ease cognition and further acquisition by users.

CHAPTER IV. DOCUMENT CONTENT ANALYSIS

Keywords: Content analysis, Text mining, Topic modelling

We will define content analysis then we will rather focus on text analysis, in presenting some text mining methods.

IV.1. Introduction to content analysis

IV.1.1. Definition

Content analysis has been defined as a systematic, replicable technique for compressing many words of text into fewer content categories based on explicit rules of coding (Berelson, 1952; Krippendorff, 1980; Weber, 1990).

(Holsti, 1969) offers a broad definition of content analysis as,

"any technique for making inferences by objectively and systematically identifying specified characteristics of messages".

Under Holsti's definition, the technique of content analysis is not restricted to the domain of textual analysis, but may be applied to other areas such as coding student drawings, or coding of actions observed in videotaped studies. In order to enable replication, however, the technique can only be applied to data that are durable in nature.

(Lasswell, 1948) formulated the core questions of content analysis: *"Who says what, to whom, why, to what extent and with what effect?"*

Thus content analysis is a systematic, objective and replicable process which provides an identification, a categorization of the manifest characteristics (e.g. source, message, recipient,…) of a document durable in nature (textual, audiovisual), in order to make inferences[6] (e.g. goal, effect,…).

[6] Inference : the reasoning involved in drawing a conclusion or making a logical judgment on the basis of circumstantial evidence and prior conclusions rather than on the basis of direct observation.

IV.1.2. Goals

Content analysis enables researchers to sift through large volumes of data with relative ease in a systematic fashion (US General Accounting Office, 1996). It can be a useful technique for allowing us to discover and describe the focus of individual, group, institutional, or social attention (Weber, 1990). It also enables inferences to be made which can then be corroborated using other methods of data collection.

IV.1.3. Requirements

According to (Krippendorff, 1980), six questions must be addressed in every content analysis:

- Which data are analysed?
- How are they defined?
- What is the population from which they are drawn?
- What is the context relative to which the data are analysed?
- What are the boundaries of the analysis?
- What is the target of the inferences?

This *"check-list"* shows that content analysis depends on the nature of the analysed corpus or media base, its structure, its users (source and target), and its environment. Moreover, it suggests that content analysis can be limited in some applications.

IV.1.4. Limits

At least three problems can occur when documents are being assembled for content analysis (US General Accounting Office, 1996).

- First, when a substantial number of documents from the population are missing, the content analysis must be abandoned.
- Second, inappropriate records (e.g., ones that do not match the definition of the document required for analysis) should be discarded, but a record should be kept of the reasons.
- Finally, some documents might match the requirements for analysis but just be uncodifiable because they contain missing passages or ambiguous content.

Moreover, (Weber, 1990) notes: *"To make valid inferences from the text, it is important that the classification procedure be reliable in the sense of being consistent: Different people should code the same text in the same way"*. The validity, inter-coder reliability and intra-coder reliability are subject to intense methodological research efforts over long years.

Document content analysis is not a miraculous process. We must understand its boundaries to increase its performances. A preliminary work must be done before on the integrity of corpus, to ensure the efficiency of the analysis and researches about inter and intra-coder reliability must be continued.

IV.1.5. Typologies

Manifest content and latent meaning

One distinction is between the manifest contents of communication and its latent meaning. *"Manifest"* describes what an author or speaker definitely has written, while latent meaning describes what an author intended to say/write.

Normally, content analysis can only be applied on manifest content; that is, the words, sentences, or texts themselves, rather than their meanings. But we can ask ourselves about a means to produce a qualitative approach, based on inferences.

Quantitative and qualitative analysis

According to Zipf's law[7], the assumption is that words and phrases mentioned most often are those reflecting important concerns in every communication. Therefore, quantitative content analysis starts with:

- word frequencies,
- space measurements,
- time counts (for radio and television time)
- keyword frequencies.

However, content analysis extends far beyond plain word counts, e.g. with Keyword In Context[8] routines words can be analysed in their specific context to be disambiguated. Synonyms and homonyms can be isolated in accordance to linguistic properties of a language.

[7] Zipf's law : named after the Harvard linguistic professor George Kingsley Zipf (1902-1950), is the observation that frequency of occurrence of some event (*P*), as a function of the rank (*i*) when the rank is determined by the above frequency of occurrence, is a power-law function $P_i \sim 1/i^{a}$ with the exponent *a* close to unity (1).

[8] Key Word In Context (KWIC) is the most common format for concordance lines. A KWIC index is formed by sorting and aligning the words within an article title to allow each word (except the stop words) in titles to be searchable alphabetically in the index. It was a useful indexing method for technical manuals before computerized full text search became common.

A further step in analysis is the distinction between quantitative approaches and qualitative approaches. That one sets up a list of categories derived from the frequency list of words and controls the distribution of words and their respective categories over the texts. While methods in quantitative content analysis in this way transform observations of found categories into quantitative statistical data, the qualitative content analysis focuses more on the intentionality and its implications. We could quote some methods, defined by (Janis, 1949):

- Pragmatic content analysis: procedure which classify signs according to their probable causes or effects (e.g., counting the number of times that something is said which is likely to have the effect of producing a specified feeling)
- Semantic content analysis: procedure which classify signs according to their meanings (e.g., counting the number of times that something is referred to, irrespective of the particular words that may be used to make reference).

Prescriptive and open analysis

(McKeone, 1995) has highlighted the difference between prescriptive analysis and open analysis. In prescriptive analysis, the context is a closely-defined set of communication parameters (e.g. specific messages, subject matter); open analysis identifies the dominant messages and subject matter within the text.

Content analysis has many different approaches, which depend on the methods (quantitative or qualitative), and the nature of corpus (only some types of files or all types of files).

If quantitative tools have been performed, content analysis has now to go towards qualitative models, and must find means to create inferences and maybe emphasize the latent meaning of documents.

IV.2. Text analysis methods

IV.2.1. Text analysis in content analysis

As said previously, content analysis is focused on any type of artefact, textual as well as audiovisual media.

If the part of video and audio is rising for the collaborative work and the diffusion of knowledge (for instance with the advent of social website like Youtube or last.fm), the textual corpora remains however the most used artefact. Moreover, although science made some progress in voice or image recognition, frequency in analysis… audiovisual analysis is still limited.

Conversely, analysis of natural language texts has known a huge improvement, mixing the progresses of computational linguistic, statistic and data mining. Two types of methods have emerged:

- the basic text mining, based on lexical, syntactic, and semantic analysis and combined with statistic and supervised data mining.
- More recently, a new text mining, called topic modelling, based on statistic and unsupervised learning data mining, and not using a linguistic approach.

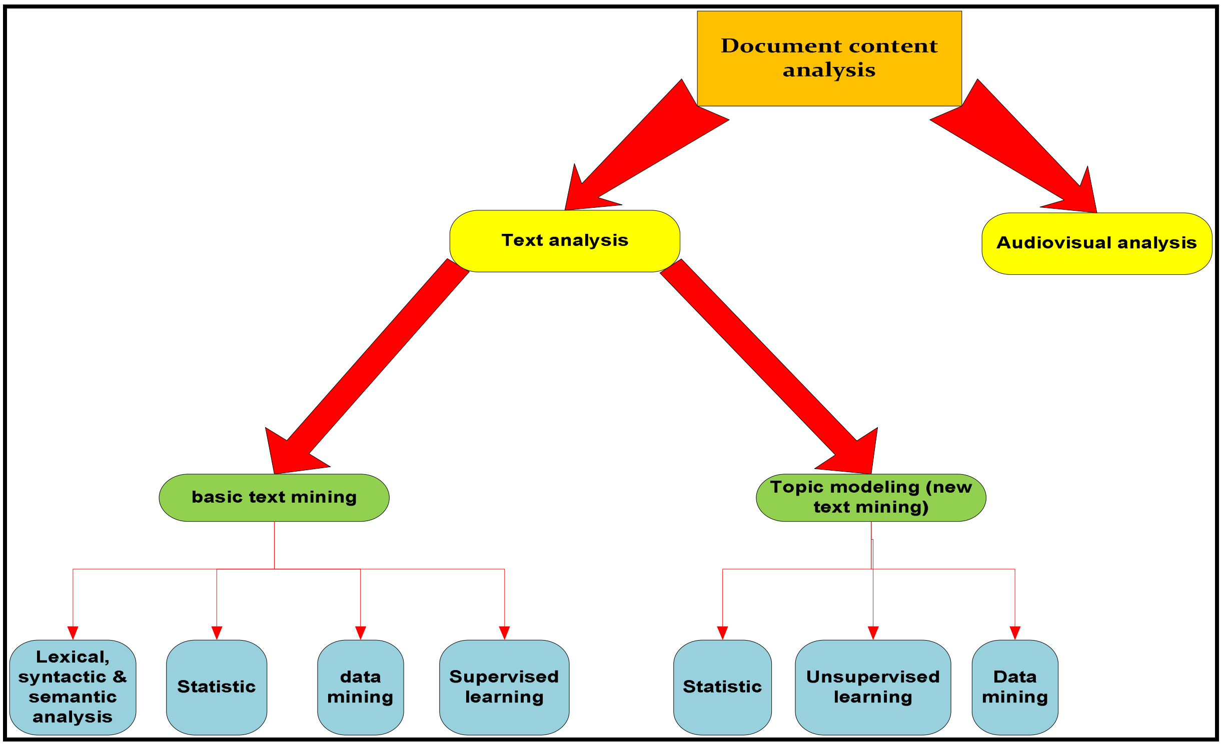


Figure 11: Text analysis methods in content analysis

We will now present deeper these two kind of text mining, in showing the possibilities and the results they supply with.

IV.2.3. Linguistic and statistical method: Usual Text Mining

<u>Definition</u>

Text mining or knowledge discovery from text (KDT) was for the first time mentioned by (Feldman, et al., 1995). It is also known as text data mining (Marti, 1999).

As the standard data mining techniques are essentially designed to operate on structured databases, text mining is described as the process of extracting interesting and non-trivial patterns or knowledge from unstructured text documents (like textual databases, word-processing files, e-mail and news collections, Web pages and other text repositories). According (Simoudis, 1996), it is often viewed as an extension of data mining or knowledge discovery from databases (KDD).

Text Mining uses a combination of statistical natural language processing[9] (NLP) and powerful mining algorithms to extract non trivial information and knowledge from these unstructured textual data (cineca, 2006).

<u>Text mining processes</u>

The figure 11 shows the decomposition of the process of text mining.

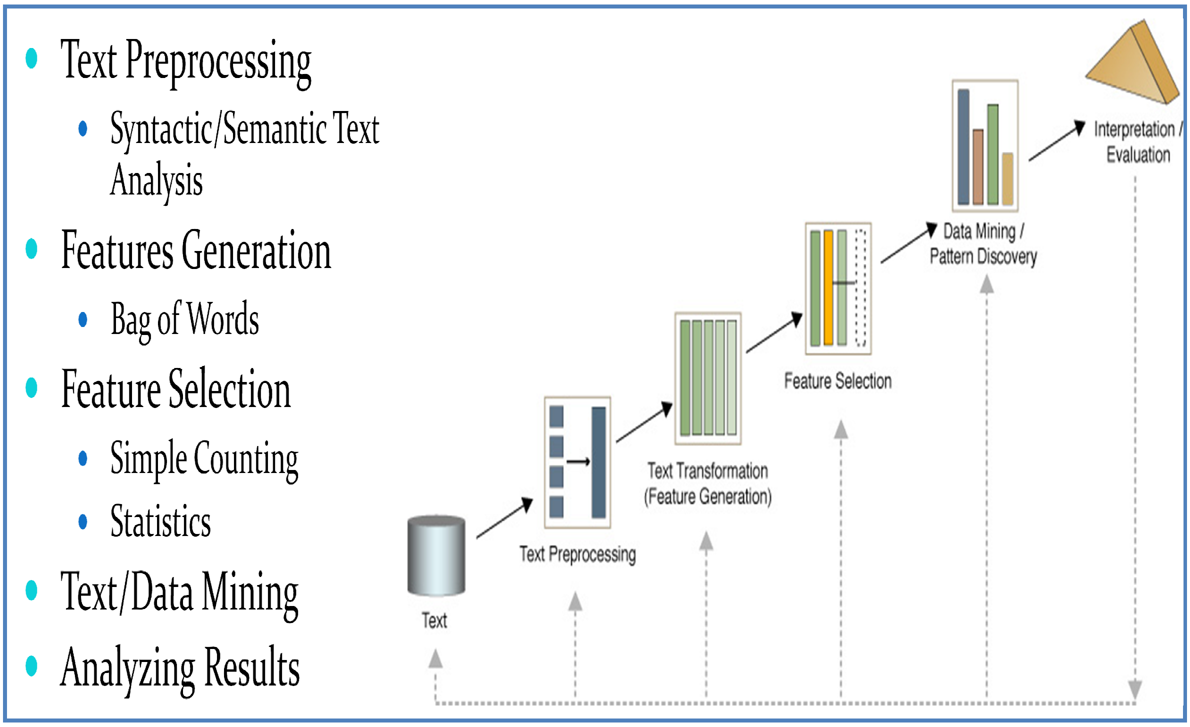


Figure 12: basic text mining process (Even-Zohar, 2002)

[9] NLP is a subfield of AI and linguistics. Language technology is often called Human Language Technology (HLT) or Natural Language Processing (NLP) and consists of computational linguistics (or CL) and speech technology as its core but includes also many application oriented aspects of them. Language technology is closely connected to computer science and general linguistics.

- First there is an application of the computational linguistic NLP:
  - in the step of text pre-processing, words, grammatical structures, pieces of sentences and meaning are analysed, with Part-of-Speech[10] (PoS) and word sense disambiguation methods, then they are parsed.
  - The step of feature generation reduces the dimensionality of pre-processed texts, with the help of some methods of Bag-of-Words[11]. Stemming[12] and lexical stop word lists[13] are used.
- Then statistic, to reduce again the dimensionality, in removing for example the words occurring in only few or in too many documents.
- Once dimensionality reduced and information structured, data mining techniques can be utilized to make inferences and discover non-trivial patterns.
- The results are analysed, in order to:
  - improve the text pre-processing (for example in adding some words in the stop-lists…), and the statistical or data-mining analysis (in defining more precisely the bounds of these analyses, for example in setting the number). The user *"teaches"* the computer to improve its performance. That shows that this basic text mining has a *"supervised learning".*
  - Feed results into information structures and organize them in visualization tools

<u>Advantages</u>

According the Italian consortium of universities in information analysis and management (cineca, 2006), *"using text mining applications, users can turn volumes of electronic documents into new insightful and valuable information about their everyday working activities. It is even more beneficial when it is used as a complementary tool to document and content management systems and other knowledge management projects"*.

Its benefits are:

- Uncover *"hidden"* content of documents including useful relationships
- Relate documents across previously unnoticed divisions, people, organisations, customers and expertise;
- Group documents by predefined common themes (supervised categorization)
- Find similar documents in content related to each other (clustering);

[10] Parts-of-Speech are lexical categories, like nouns, verbs, adverbs, adjectives… PoS methods focus on the identification of words in these lexical categories
[11] consideration of each word and its occurrence, the order of words is not important
[12] identification of word by its roots
[13] alienation of noise, for example the words "the", "an", "but"…

IV.2.3. Non linguistic and unsupervised method: Topic Modelling

Definition

Topic modelling is a new kind of text-mining technique, which is based on the idea that individual documents are made up of one or more topics. It uses emerging technologies in computer science to automatically cluster topically similar documents by determining the groups of words that tend to co-occur in them (Block, 2006).

Topic modelling is based on a statistical method, LDA (Latent Dirichlet Allocation), which gives a stochastic matrix representing the occurrences by thematic: This allows the characterization of the whole corpus. It can also associate a document to a topic with a probability.

Non linguistic and unsupervised learning

Moreover, it is independent from knowledge about language, unlike the dictionary based methods and old technique. It doesn't need to know the grammatical rules for instance.

This process has none prior knowledge about the topics of the documents (Rigouste, et al., 2006). So topic modelling is unsupervised learning, unlike the older text mining methods.

Indeed, older text-mining techniques require the user to come up with an appropriate set of topic categories and manually find hundreds to thousands of example documents for each category. This human-intensive process is called supervised learning. In contrast, topic modelling, a type of unsupervised learning, doesn't need suggestions for an appropriate set of topic categories or human-found example documents. This makes retrieving information easier and quicker (UCI, 2006).

Topic modelling is a new text-mining and an unsupervised-learning technique, which identifies groups of co-occurring words, it is to say topics, and enables to categorize and cluster topically similar documents, without knowledge about languages.

Examples of topic modelling application

This method can for instance, according (Block, 2006):

<table>
<tr><th>Possibilities</th><th>Examples</th><th>Illustrations</th></tr>
<tr><td>Figure out common words of a corpus of various documents</td><td></td><td></td></tr>
<tr><td>Classify and “makes interpretable” topics, by sorting and categorizing relevant words</td><td>- For instance, the topic model might group together the following words as those most likely to appear in a particular subset of documents: Indian fort men town party off killed people came letter day French… In this case, we can easily identify that list as a topic related to interactions between colonists and Native Americans—perhaps we might label it INDIANS (Block, 2006).</td><td></td></tr>
<tr><td>Link topics and words</td><td>- Topic modelling can also show users the most likely topics associated with particular words—type a word into a search box, and you can get a list of the most likely topics in which that word appears (Block, 2006).</td><td><table>
<tr><th>Word</th><th>Topic 1</th><th>Topic 2</th><th>Topic 3</th><th>Topic 4</th></tr>
<tr><td>Cherokee</td><td>Indian (100%)</td><td></td><td></td><td></td></tr>
<tr><td>Negro</td><td>Servant/Slave (96%)</td><td>Disaster (4%)</td><td></td><td></td></tr>
<tr><td>Woman</td><td>Servant/Slave (82%)</td><td>Disaster (10%)</td><td>Indian (4%)</td><td>Runaway (3%)</td></tr>
</table></td></tr>
<tr><td>Link topics and documents</td><td>- Because each set of topic words can be linked to the documents that most highly correlate to that topic, users can find individual documents on those topical subjects. Those documents that most exclusively focus on a topic are that topic’s most highly ranked (Block, 2006). So, topic modelling get out the top-ranked articles related to a specific topic.

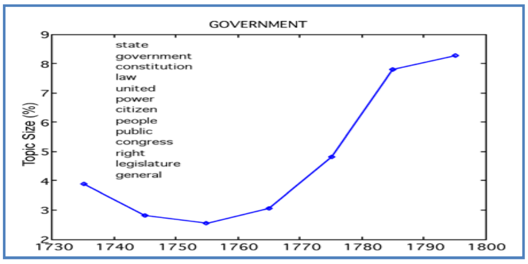
- The topic model allows users to see the multiple topics that a document simultaneously contains.</td><td><table>
<tr><th>% of article</th><th>Most likely words in topic in order of likelihood</th><th>Human-added topic label</th></tr>
<tr><td>47</td><td>court assembly general county office law election judge year council city person justice...</td><td>LAW & COURTS</td></tr>
<tr><td>18</td><td>act person aforesaid within authority further thereof enacted hereby officer state...</td><td>LEGISLATION</td></tr>
<tr><td>9</td><td>right great people power law colony act without britain subject country America liberty...</td><td>POLIT IDEOLOGY</td></tr>
<tr><td>8</td><td>state government constitution law united power citizen people public congress...</td><td>GOVERNMENT</td></tr>
</table></td></tr>
<tr><td>Track topics over times</td><td>- Topic modelling can also chart the changing prevalence of each topic over time. Not surprisingly, a topic related to the kinds of political issues discussed at the founding of the United States (state government constitution law united power citizen people public congress right legislature…) increased in prevalence when it is supposed to: in the Revolutionary and early national eras.</td><td>
</td></tr>
<tr><td>Link topics and people/departments</td><td>- Topic model can link also topics to people and department, in analysing documents and authorships. Thus, the method can characterize and categorize persons and/or laboratories for example. It is a good means to identify skills and facilitate collaborative work and creation of partnerships.</td><td></td></tr>
</table>

Table 2: examples of possibilities of Topic modelling

IV.3. Conclusion

Content analysis is a replicable process identifying, characterizing and categorizing every type of media. Nevertheless, the scientific progress has focused up to now more on textual source, easier codifiable and analysable.

As we have just seen, text mining methods are powerful tools, which support Many processes of classification (by identifying and categorizing topics) and retrieval of people and information (by linking topics to documents, words, people…).

Two types of text mining exist:

- a classic text mining, based on an approach mixing linguistic (NLP), statistic and data mining algorithms,
- another recent method, topic modelling, which uses solely a powerful statistic tool (LDA) with data mining algorithms but without linguistic analysis.

Thus the first step of text pre-processing in previous text mining is suppressed in the new technique, and the loop of supervised learning is also disappearing, automating more analysis and extraction of content in documents.

After having observed content analysis and its methods, we will shed the light on the methods or tools for organizing extracted information in structures enabling the creation of inferences, semantic and logic links.

CHAPTER V. INFORMATION ARCHITECTURE WITH "KNOWLEDGE MAPS"

Keywords: knowledge map, ontology, Topic map, conceptual map

In this chapter I will present a type of information structure, knowledge map, which enables to organize knowledge stored in artefacts or held by people. After a short listing of different knowledge map, I emphasize three of them, in attempting to point out their advantages and their constraints.

V.1. Introduction to knowledge maps

V.1.1. Definition

"Knowledge" maps are information structures which work like yellow-pages that contain a *"who knows what"* list.

A knowledge map does not store knowledge (Baroni de Carvalho, et al., 2002) but allows accessing knowledge held by people, facilitating *"the development of interpersonal connections around topics of interest*" (Hertzum, et al., 2000).

It provides an expert locator feature that helps users find the best-suited experts to work on a specific problem or project.

(Hertzum, et al., 2000) proposes two approaches for supporting searches for people:

- "to extend document retrieval systems by explicitly exploiting the fact that documents tell a lot about the work activities of their authors and thereby provide a rich description of the authors' experience and competencies"
- "to develop models for classifying people's expertise" (without eliciting people's expertise).

These approaches are presented as "the ask a program/document" and "ask a person" paradigms into information seeking (Yiman Seid, et al., 2003).

V.1.2. Classification

According to (Greenwood, et al., 2006), we can distinguish many different tools and techniques to organize knowledge[14]:

1. Concept map 2. Mind Map /Idea map 3. Concept circle diagram 4. Semantic map 5. Cognitive map 6. Process map 7. Social mess map / Cross boundary causality map 8. Conceptual map	9. Knowledge flow map 10. Causal map 11. Ontology 12. Petri net 13. Cluster Vee diagram 14. Thesauri 15. Visual thinking network 16. Topic map 17. Perceptual map

Table 3: Typologies of Knowledge Maps

We are going now to define more precisely three types of structures of knowledge maps, conceptual map, topic map and ontology, which are characterized by a computer language and enable to structure concepts and the links between concepts.

V.2. Knowledge map structures

V.2.1. Conceptual maps, an informal means to organize concepts

Definition

To understand conceptual maps, we need to give first two notions defined by (Novak, et al., 2006):

- concept as “a perceived regularity in events or objects, or records of events or objects, designated by a label”. The label of a concept is usually a word.
- propositions are “statements about some object or event in the universe, either naturally occurring or constructed. Propositions contain two or more concepts connected with other words to form a meaningful statement” They are also called “semantic units”.

That leads to the definition:

Conceptual maps are artefacts for organising and representing knowledge, by drawing relations between concepts in the form of propositions (Novak, et al., 2006).

[14] see appendix 3 for more details

Moreover, conceptual maps are structured in a hierarchical way, where the most general concepts lie in the root of the tree and, and as we descend the structure, we find the more specific ones.

Roles

These maps aimed at understanding the changes in time of the knowledge that children had of science.

When they are well designed, taking into account the context and motivation of their audience, they constitute a teaching and a learning instrument that facilitates understanding and assimilation of the concepts and their relations.

Although their origin is to learning, their application to Information Visualisation configures them as useful tools to convey complex messages in a clear way.

Patterns

Conceptual maps include (Dürsteler, 2004):

- concepts,
- relationships between concepts defined by linking words or linking phrases.
- cross-links, which are relationships or links between concepts in different segments or domains of the concept map. Cross-links help us see how a concept in one domain of knowledge represented on the map is related to a concept in another domain shown on the map. In the creation of new knowledge, cross-links often represent creative leaps on the part of the knowledge producer. [15]
- sometimes specific examples of events or objects that help to clarify the meaning of a given concept.

We are going now to apprehend the ontology map tools, which give a more formal background to represent knowledge.

V.2.2. Ontology, a formal architecture

Definition

Many definitions of ontology have been proposed in literature.

In computer science, ontology is the attempt to formulate an exhaustive and rigorous conceptual schema within a given domain, typically a data structure containing all the relevant entities and their relationships and rules.

[15] There are two features of concept maps that are important in the facilitation of creative thinking: the hierarchical structure that is represented in a good map and the ability to search for and characterize new cross-links.

According to (W3C, 2004), ontology defines the terms used to describe and represent an area of knowledge. Ontology is used by people, databases, and applications that need to share domain information, where a domain is just a specific subject area or a wider area of knowledge, like tool manufacturing. Ontology includes computer-usable (computable) definitions of basic concepts in the domain and the relationships among them.

From all these definitions, we can keep one, which is proposed by (Gruber, 1995):

Ontology is a formal, explicit specification of a shared conceptualization.

- A “conceptualization” refers to an abstract model of some phenomenon in the world, which identifies the relevant concepts of that phenomenon.
- 'Explicit' means that the type of concepts used and the constraints on their use are explicitly defined.
- “Formal” refers to the fact that the ontology should be machine understandable.
- “Shared” reflects the notion that ontology captures consensual knowledge, that is, it is not restricted to the knowledge view of some individual, but reflects a more general view shared and accepted by a group.

Roles

According (Werthner, et al., 2003) the role of ontology is twofold:

- they support human understanding and communication
  - They allow to identify and define unambiguously the key concepts and the relevant terms to a given domain. Therefore, the use and exchange of data, information, and knowledge among people and organizations is facilitated.
  - Moreover, ontology facilitates the integration of different user perspectives, while capturing key distinctions in a given perspective.
  - Furthermore, the use of ontology enables the cooperation among people at different levels: internal cooperation, external cooperation, and integrated cooperation (external cooperation where internal knowledge is shared to solve a complex task)
- they, in machine-processable form, facilitate content-based access, communication and integration across different information systems. They can be used at the following three levels:
  - Design and development of software systems. At this level ontology plays an important role in the specification, reliability, and reusability of software systems.

- At the Communication level, ontology facilitates the data exchange: among system designers fostering mutual understanding; among and among the different software tools and application systems fostering reconciliation.
- At the Interoperability level, Ontology Based services support different software systems to cooperate at different levels: Data Interoperability, Function Interoperability, and Process Interoperability.

These roles are both achieved by explicating and formalizing the meaning, or semantics, of organization and enterprise application information resources. So ontology is powerful for inference making and addition of logic to content.

Patterns

Knowledge in ontology is mainly formalized using five kinds of components:

- A class or concept represents a set of entities within a domain. The classes in the ontology are usually organised in taxonomies.
- Relations represent the interaction between concepts of the domain. The relations can be organised in taxonomies.
- Functions are a special case of relations in which the n-th element of the relationship is unique for the n-1 preceding elements.
- Axioms are used to model sentences that are always true. They can be used in ontology to constrain values of classes, to define the arguments of relations etc.
- Instances are used to represent specific individual elements.

Ontology is a good structure for organizing topic and making inferences. But it is too highly formal and has many constraints. Let us now study the topic maps.

V.2.3. Topic maps, semi-formalized tools

Definition

(ISO/IEC13250, 1999) defines Topic Maps as:

A topic map defines a multidimensional topic space — a space in which the locations are topics. The distance between topics are measurable in terms of the number of intervening topics which must be visited in order to get from one topic to another, and the kinds of relationships that define the path from one topic to another, if any, through the intervening topics, if any.

Several topic maps can provide topical structure information about the same information resources. The Topic Maps architecture is designed to facilitate merging topic maps without requiring the merged topic maps to be copied or modified.

Because of their extrinsic character, topic maps can be thought of as overlays on, or extensions to, sets of information objects.

Roles

Topic maps enable multiple, concurrent views of sets of information objects. The structural nature of these views is unconstrained; they may reflect an object oriented approach, or they may be relational, hierarchical, ordered, unordered, or any combination of the foregoing. Moreover, an unlimited number of topic maps may be overlaid on a given set of information resources.

Topic maps can be used:

- To qualify the content and/or data contained in information objects as topics to enable navigational tools such as indexes, cross-references, citation systems, or glossaries.
- To link topics together in such a way as to enable navigation between them. This capability can be used for virtual document assembly, and for creating thesaurus-like interfaces to corpora, knowledge bases, etc.
- To filter information set to create views adapted to specific users or purposes. For example, such filtering can aid in the management of multilingual documents, management of access modes depending on security criteria, delivery of partial views depending on user profiles and/ or knowledge domains, etc.
- To structure unstructured information objects, or to facilitate the creation of topic-oriented user interfaces that provide the effect of merging unstructured information bases with structured ones. The overlay mechanism of topic maps can be considered as a kind of *external mark-up mechanism*, in the sense that an arbitrary structure is imposed on the information without altering its original form.

Patterns

. In general, the structural information conveyed by topic maps includes:

- Occurrences: they connect the topics to information resources that contain information about them (by gathering addressable information objects around topics with URIs).
- Relationships between topics (‘associations’).[16]

[16] Two topics may be connected through an association, and they can also be connected by virtue of sharing an occurrence.

In addition, information objects can have properties, as well as values for those properties, assigned to them externally. These properties are called *facet types*.[17]

<u>Topic map and Ontology</u>

When we look closely at any topic map, we will find classes of topics, association types, role types, occurrence types, implicit rules of cardinality, etc. All these symptoms show the presence of some underlying, implicit if not explicit, ontology.

Thus, Topic maps have been designed to be deliberately 'ontology-agnostic', in the sense that they are intended to be able to represent and manage any kind of subjects and relationships, in an ontological context (Vatant, 2003).

V.3. Conclusion

We have studied three main types of knowledge maps, providing an information architecture easing the future access to people and knowledge.

Concept map is a user-friendly tool, oriented visualization, but it is not formal, which avoid computers “understanding” and thus automating inferences.

Ontology is a formal background to create and automate logic and infer relationships between concepts, but its shared and common language and the accurate definition of its classes or rules is very reluctant.

At long last, topic maps combine seemingly conceptual maps and ontology structures. Indeed, TM uses the formalism and the “inference-making” capacity of ontology, with the user-friendly simplicity of conceptual maps. Moreover, it links easily document resources to topics.

TMs offer semi-formalized means to structure information, in making inferences, easing the human use and linking resources to concepts. It combines the ability of being understood by human and computer.

[17] The word facet can mean one side of a many-sided, polished object, or one segment of a compound eye (e.g. an insect's). Its metaphorical use here captures the idea that a facet is a property of a set of information objects that can be used to create a view of them.

To conclude, we are going to summarize the different advantages and limits of the three systems in the table below:

Knowledge map structures	**Advantages**	**Limits**
Concept Map → human understanding	- User-friendly tool to organize information. Because it is not standardized, natural language can be used. - In addition to representation of concepts and relationships, possibilities to add cross-links and occurrences are interesting. - The hierarchy representation is also a good point to conceptualize better specific relations, e.g. membership relationships.	- Their informal language, which prevents computers from “understanding” and making inferences - The absence of classes and rules above concepts and relationships
Ontology → computer understanding	- Communication between computer, with its formal languages - Definition of rules and classes, which categorize and level concepts and relations, and can organize semi-automatically knowledge (e.g. rules can define relationships between concepts of specific classes). - Hierarchy, with a taxonomy model and the possibility to define equivalence classes	- Shared controlled vocabulary and formal language prevent from using natural language - It is not very user-friendly to represent knowledge.
Topic Map → human and computer understanding	- They have an ontological frame. We can define classes and rules, to make inferences be possible. - They let people use a not controlled language. - They have a standardized computer language, which is “understandable” by computer. - They can link information resources to instances with URIs.	

Table 4: Comparison of advantages and limits of the different information architectures

After have defined tools which could give means to analyse and organize document content in topics and concepts, we will now describe tools for represent them.

CHAPTER VI. VISUALISATION TOOLS

Keywords: complexity, context, dynamics, graph, tree, map

After stressing the interest and roles of information visualization, we will study and compare the different methods and tools to support it.

VI.1. Interest and roles of information visualization

Today's media technology provides a framework in which knowledge can be archived and transmitted, but current systems of accessing, organizing and navigating information are proving insufficient. One strategy to make more sense of a complex information space is information visualization, the visual presentation on an interactive map. If artefacts, like texts or pictures, allow capture of meaning and make explicit knowledge, digital media gives means to organize and simplify domain knowledge. **Both of them are complementary: artefacts "contain" knowledge and maps locate and situate knowledge.**

According to (Judelman, 2004), "*visualization takes advantage of visual and spatial cognitive powers to reduce the cognitive effort required for processing complex information".* The mapping of data parameters to location, colour, or form produces images which can reveal objects, patterns and relationships which remain undetectable when presented as lists or tables. (Judelman, 2004) develops further these advantages, in decomposing the roles of visualization in three dimensions:

- In order to understand and reduce the complexity, "complexity spaces" should show the topology, the hierarchy ("classify") or paths ("route") of information architecture
- So as to situate knowledge and reduce complexity, "context spaces" tends to show the semantic relationships ("chart") or content of information ("explore"), pointing out the similarity or difference.
- Eventually, "dynamic spaces" can be used to visualize the spatiotemporal changes ("evolve" temporally or "flow" spatially) in information or knowledge.

These three dimensions are summarized in the following scheme:

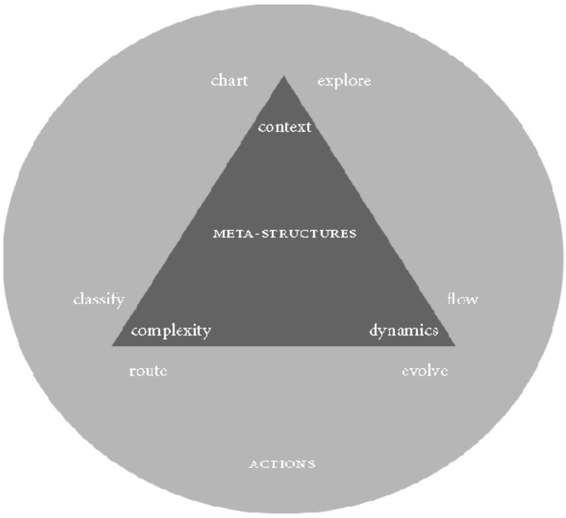


Figure 13: roles and power of visualization (Judelman, 2004)

(Card, 1999) outlines another aspect known as "focus and context". This refers to the simultaneous overview (context) of the entire information structure with a detailed view (focus) of a particular section.

Finally, (Amende, et al., 2004) contends that *"good representation helps users to find interesting topics. An efficient navigation is important for quick access to the topic of interest"*.

Thus, information visualization is a performing means to capture meanings and create inferences by visualizing relationships, changes or context. Indeed, it facilitates global and focused views on a domain, reduces cognitive efforts from users and provides dynamic comprehension with interactive navigation.

VI.2. Visualisation tools

With a concept map, ontology or topic map we can structure essential information adapted to the users needs. Therefore it is important to create a comfortable visualization where users get an overview or a filtered view of topics an understanding of their changes and their associations, so that they find the information they need.

In the literature, graphs, trees and maps and the combination of them are the main diffused techniques among all the visualization methods providing easy cognition for data, information, concept, strategy… [18]

[18] See Appendix 5

VI.2.1.Graph and Tree tools

Graphs are networks of nodes and edges. Nodes represent concepts and edges the associations between the concepts. Static graph visualization shows all nodes with their associations. To avoid clutter and complexity, dynamic graph visualization displays only a limited scope of nodes and associations starting from the topic of interest and its related topics (Ahmed, 2000).

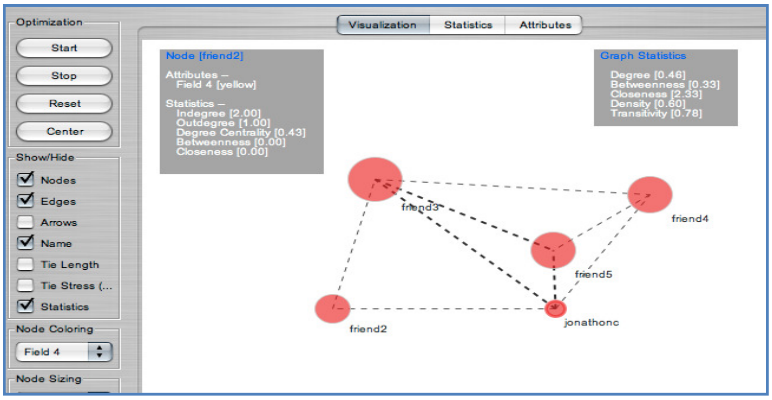


Figure 14: Example of graphs (Netvis, 2007)

Trees arrange the topics and edges in a hierarchical structure, making it easier for users to interpret (Le Grand, et al., 2003). In this way information can be better structured. Trees are often used to visualize organization structures, computer file systems, interlinked Web hierarchies and communication hierarchies (Rohrer, et al., 1997). Hyper-linked trees (site maps) guide a visitor through a web site using hyper-links between nodes, which represent a structured form of the content list referent (Oliveira, 2000). Like graphs, they can be dynamic to reduce complexity.

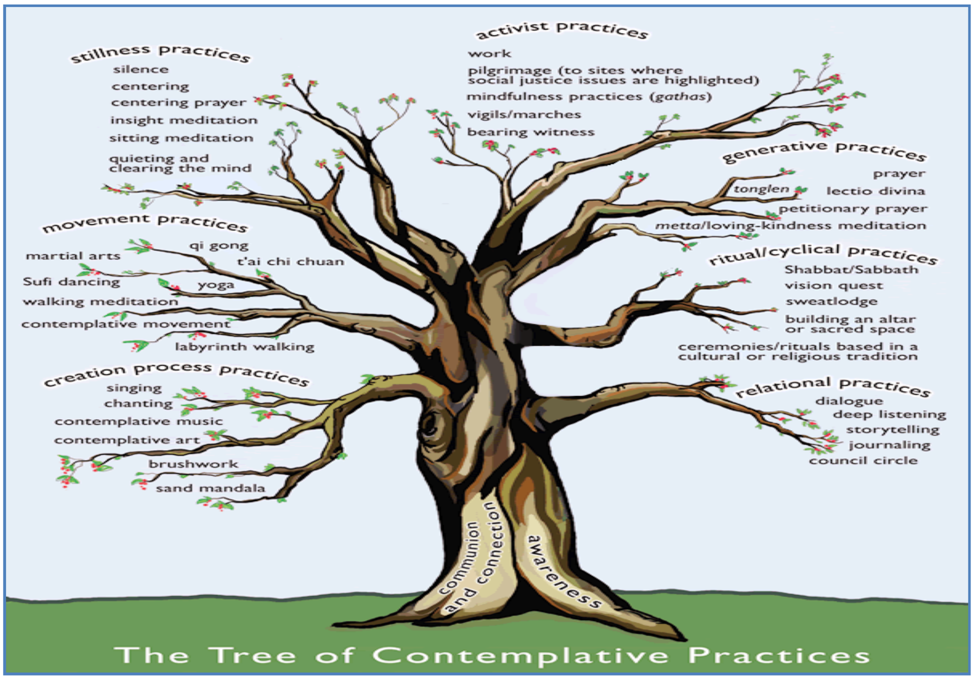


Figure 15: example of tree (contemplativemind, 2007)

VI.2.2. Map tools

Maps are arranged at a certain position on a 2- or 3-dimensional grid. Map visualisation adds some topological[19] information, and can use colour, distance or size to provide more information.

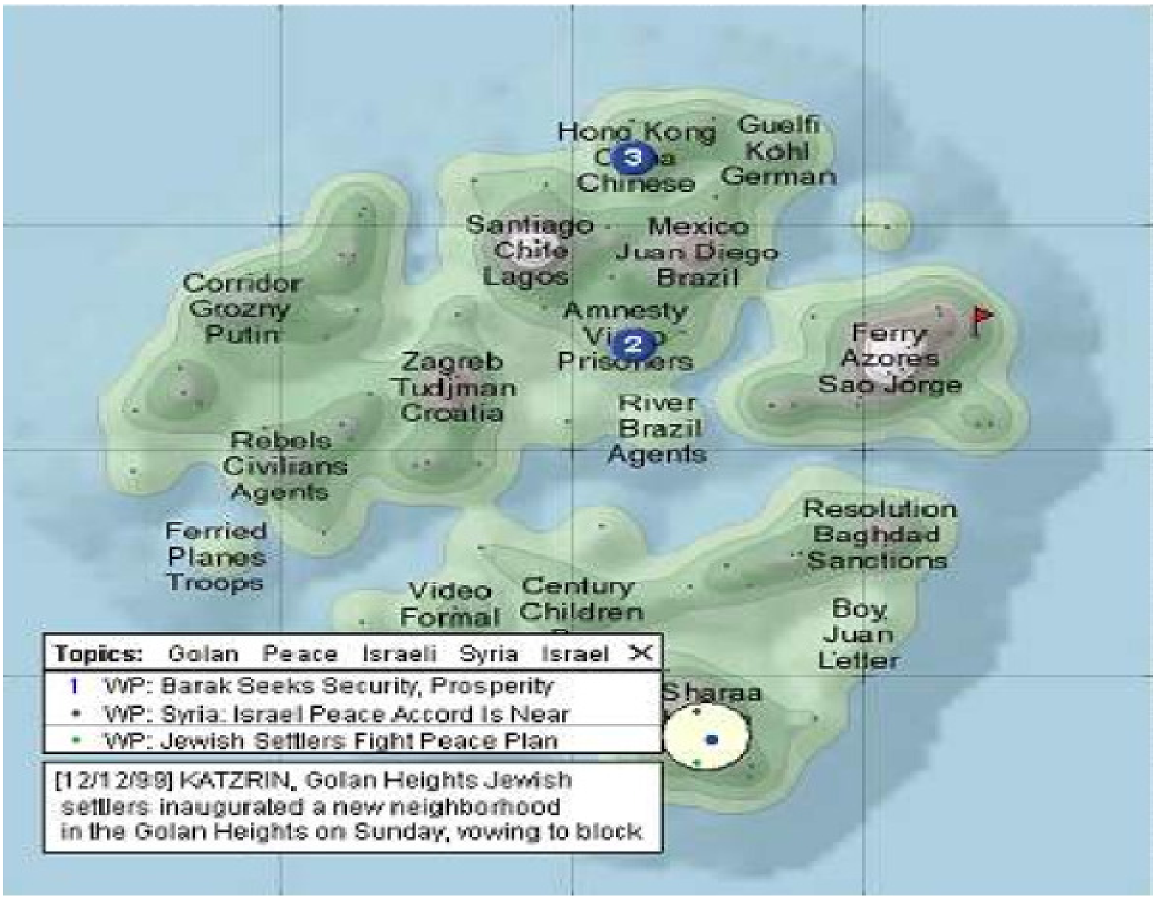


Figure 16: Example of map using topography, colour, size and distance (Dodge, 2000)

For example, it can be arranged like an original topological map. Mountains display topics, whereas related mountains (topics) are placed close to each other. The mountain's height is depending on the degree of closely related documents (occurrences) to one topic. The valleys between mountains can be interesting, because they contain fewer documents and more unique content. Avoiding complexity labels reflect only the biggest mountains on the map (Le Grand, et al., 2003).

[19] A topological map is simply a mapping that preserves neighbourhood relations.

VI.2.3. Map and tree combinations

Some other tools are hybrid, combining some aspects of both tree and map (for instance the hierarchy from trees, with topological information of maps). We will explore three types of these combinations: cluster maps, fractal maps and tree maps.

<u>Cluster maps</u>

Cluster maps are graphs gathering similar information in clusters. So as to visualize similarities and differences, they can use some map characteristics, like for instance distance (the smaller it is, the bigger the similarity rises) or colour (to emphasize the belonging to a group). A more detailed study on this system is found in (Chen, 2003).

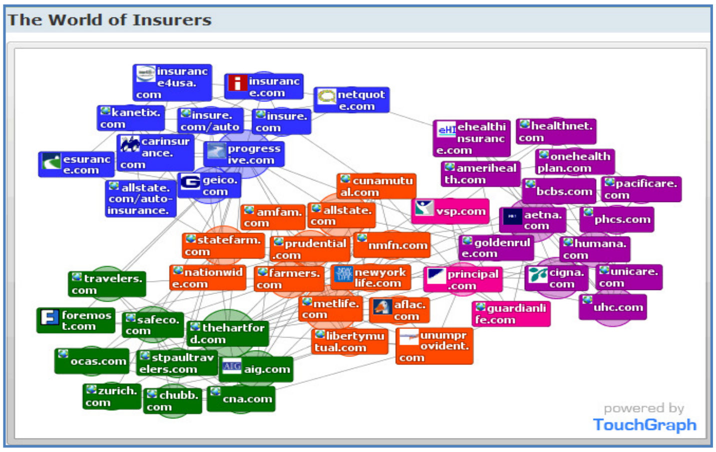


Figure 17: Example of cluster map on the insurance world (TouchGraph, 2007)

<u>Fractal maps</u>

Fractal view is an information reduction approach and an approximation mechanism to abstract complex objects and controls the amount of information to be displayed with a scale (threshold) set by users [17]. It gives some details near the focus point and important landmarks which can be further away explored by zooming. This information is flexible to the interest of users (Koike, 1995).

In some extent, it is such a view from above of a tree, using clustering aspects. Users can course the hierarchy in opening the next low levels.

Figure 18: Example of fractal map (Grokker, 2007)

Tree maps

Tree map is an invention of the Human-computer interaction laboratory at the University of Maryland (HCIL, 2007).

This system flattens node-link tree diagrams onto a 2D map filled by squares. Directory levels are contained in these squares, with subdirectories iteratively contained therein. It uses the size and the colour of the square to show the type of information, the depth in the hierarchy... The visualization is dynamic and can change according the criteria and the threshold chosen for the parameters (size, colour…).

In a way, it is a mapping representation of multi-criteria graphs.

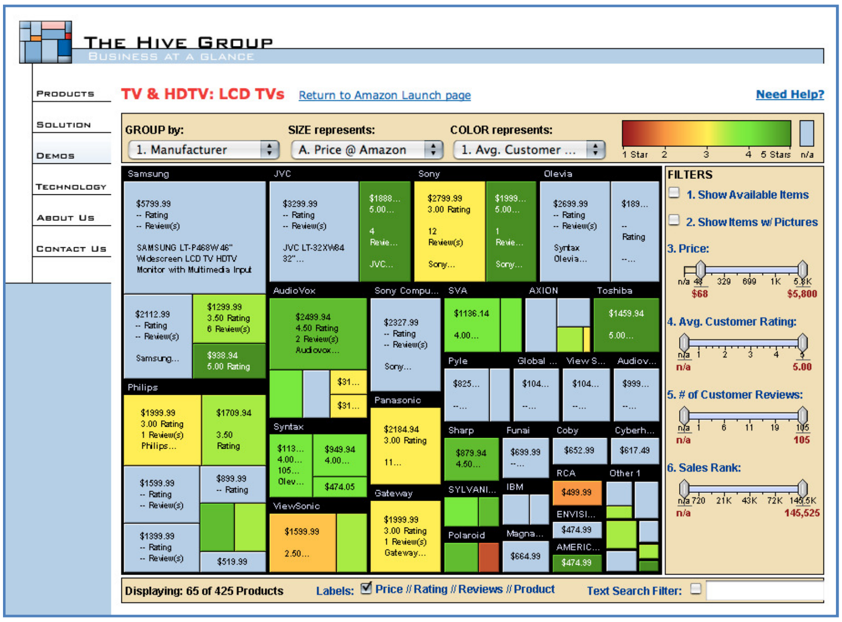


Figure 19: Example of Tree map on television choices (The Hive Group, 2007)

VI.3. Comparisons of tools

We are going to compare the visualization tools presented above. To do this we will take as criteria the two first dimensions[20] of (Judelman, 2004) and the "double view" of (Card, 1999), and the duality navigation/representation of (Amende, et al., 2004).

Thus the criteria will be:

- The display and the "reduction of complexity" (or "representation" of Amende),
- The visualization of the "context" (Card's "overview"), showing hierarchy and neighbourhood,
- A detailed view ("focus") adapted to user's needs
- A "navigation" in information

[20] We have deliberately suppressed the "dynamics" dimension of Judelman because it is not focused on the visualization of knowledge but rather on its life (its history or its moves).

Overview and navigation

(Amende, et al., 2004) contend that *"graphs and trees techniques concentrate on navigation through hyperlinks whereas maps or landscape maps concentrate on topics representation".*

Graphs and trees cannot generally visualize an overview of a domain. It is obvious that over a certain amount of information[21], complexity avoid users having a general view. By contrast, using graphs or trees, users will benefit from a good navigation between the labels.

By nature, overview is possible for all the other maps, but because generally standard maps have no paths, user cannot navigate through different concepts.

Finally, the combination of trees and maps make overview and navigation possible together. Indeed:

- cluster maps use the architecture of trees for navigation and clustering algorithms to represent group of topics,
- Fractal maps has an overview structure but enables navigation in opening lower or upper directories
- Tree maps, as their name expresses, combine overview of maps but represents also the structure of tree, each folder containing its subfolders.

Complexity reduction

Complexity reduction is finely linked to the possibility of overview. By nature, all maps, and so combinations of trees and maps, provide means to reduce complexity.

For graphs and trees, this reduction depends on their dynamics. Static tools are obviously very complex, but in dynamics graphs and trees, complexity decreases, since you can limit the scope of displayed concepts.

Focused view

For graphs, tree, maps and cluster maps, focused view is possible when the tools are dynamic. To do this, methods of limitation of scope or zooming are used.

Fractal maps and tree maps are always dynamics, so they provide always a detailed view.

[21] even for the dynamic tools, where mass of information in information in a same level or in a same degree of neighbourhood can be huge

VI.4. Map automation

In order to automate this visualisation, some algorithms were designed, to represent directly the data extracted by content analysis and structured by information architectures.

We could mention a tremendous technique, the Self-Organizing Maps (SOM) Algorithm (Kohonen, 2001). Actually it is a clustering method which is used to calculate optimal coordinates for the topics on a map.

Kohonen's SOMs are a type of unsupervised learning. The goal is to discover some underlying structure of the data.

They are also called a topology-preserving map because there is a topological structure imposed on the nodes in the network (Giraudel, et al., 2001).

VI.5. Conclusion

In this chapter, we have seen the different visualization techniques, emphasized their advantages and mentioned the possibility of creating representation automatically with SOM algorithms.

We summarized our previous observations in the following table:

	Complexity reduction	**Context/ overview**	**Detail/ focus**	**Navigation**
Graphs	-*	No	-* (limited scope)	Yes
Trees	-*	No	-* (limited scope)	Yes
Maps	Yes	Yes	-* (zooming)	No
Cluster maps	Yes	Yes	-* (limited scope or zooming)	Yes
Fractal maps	Yes	Yes	Yes	Yes
Tree maps	Yes	Yes	Yes	Yes
*: it depends whether graphs or trees are dynamic or static.				

Table 5: Comparison of visualization tools

We have also led a survey on visualization tools with these criteria, in appendix 6.

We have studied the different tools to help users for automating reification process, especially in the steps of extraction, organization and visualization. In the following section we will observe the means to enrich and assess this automatic information generation.

Section 2: Enrichment & assessment

Like in many systems or processes, automated reification must be controlled and have a feedback.

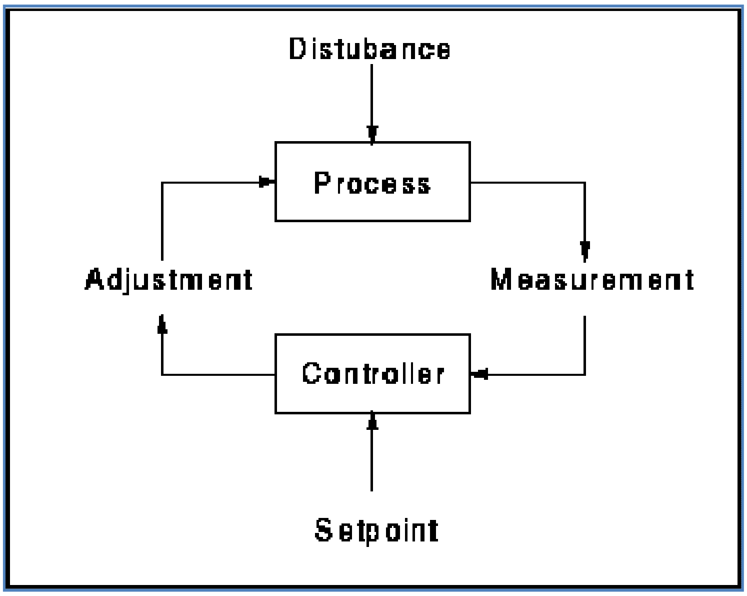


Figure 20: Feedback and control in automation (Edinburgh University, 2007)

That means that the extracted and organized content coming from reification must be assessed ("measurement") and corrected ("adjustment"), by human or machine.

In this section, we will study the different possibilities for enriching and assessing the reification process and its components. We will observe particularly the complementariness of enrichment and assessment.

Then we will focus on some tools of information filtering and their abilities for support these complementary notions.

CHAPTER VII. THE LINK BETWEEN MANUAL ENRICHMENT & ASSESSMENT

Keywords: enrichment, control, assessment, performance, indicators

We will start by studying the possibilities of actions by users, and then we will define performance, assessment and their meaning in a virtual CoP. Finally we will show the link existing between enrichment and assessment.

VII.1. Enrichment and control of information by users

VII.1.1. Roles and needs

If automation pushes information towards users, users must check if the extracted and organized information is exact and relevant and he must be able to correct it in some cases.

Furthermore, automated information comes from corpora, i.e. "static sources". Nevertheless, in order to model all the knowledge and add dynamics, users must participate and they must be able to interact with automation, in adding some information.

So it is important that information can be:

- Modified (correction),
- Completed (addition),
- Explained (simplification),
- Commented (recommendation),
- Assessed (measurement),

by users, either on Knowledge Base or visualisation tools.

VII.1.2. Actions

Users have many possibilities to interact with knowledge base or maps. We identified some usual actions for creation as below:

- Erase (an input in map or a sentence in a text for instance)
  → User corrects in deleting wrong information
- Edit (map or texts)
  → User modifies or adds some information
- comment or annotate texts
  → User gives (adds) his advice (personal recommendation and explanation) on information
- add keywords on texts
  → User provides (adds) some metadata (recommendation) defining information (explanation) held by a text
- summarize texts
  → User sum up the content of a text
- rate texts
  → User gives his advice with a score, i.e. manually assess items

We summarize these actions linked to the needs identified above in the following table:

Action/needs	Correction	Addition	Explanation	Recommendation	Measurement	Used resources
Erase	X					Map, texts
Edit	X	X				Map, texts
Comment		X	X	X		texts
Add keyword		X	X	X		texts
Summarize		X	X	X		texts
Rate		X	X	X	X	texts

Table 6: Actions/needs matrix for enrichment and control of information by users

VII.1.3. Existing solutions

Definition of metadata

We have found some definitions emphasizing the possibilities of metadata for identification, assessment, enrichment and location.

"Metadata is structured, encoded data that describe characteristics of information-bearing entities to aid in the identification, discovery, assessment, and management of the described entities." (American Library Association, 1999)

"[Metadata is a set of] optional structured descriptions that are publicly available to explicitly assist in locating objects." (Bultermann, 2004)

To keep a simpler definition (given by wikipedia), **metadata is a data about data - more specifically information (data) about a particular content (data).**

For instance, the context of a library, where the data is the content of the titles stocked, metadata about a title might typically include a description of the content, the author, the publication date and the physical location.

So all the additive but not directly corrective actions (comment, put keywords, summarize or rate) imply the use of metadata.

Tools and examples

Many systems exist and help users in their actions and in *"metadata feeding*". We have enlisted some tools and sorted them by their inputs, words, paragraphs, scores:

- Tags, keywords (words)
  - Some practices on internet use chosen keywords called tags. Some of them are named folksonomies, a contraction of folks (friends) and taxonomy (classification). They describe processes of collaborative categorization (Palmer, 2006). Good examples of folksonomy can be found on http://del.icio.us or http://www.last.fm.

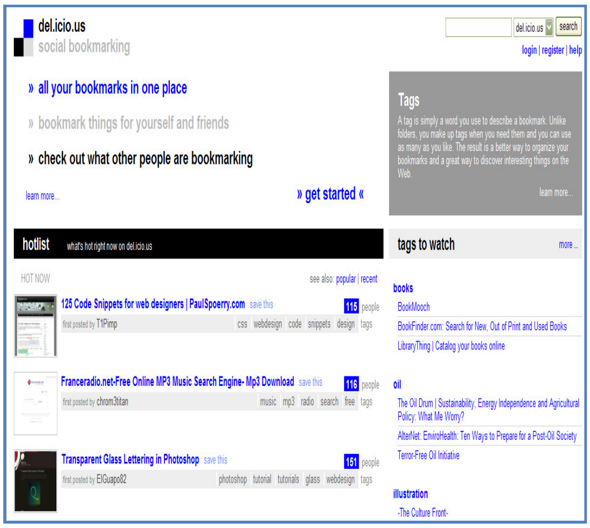


Figure 21: del.icio.us, a folksonomy website

- There are also some other systems based partially on content analysis, which suggest keywords to users. The users are finally the last to decide and choose the relevant keywords. One instance of such systems is the metadata system of EDEN™.

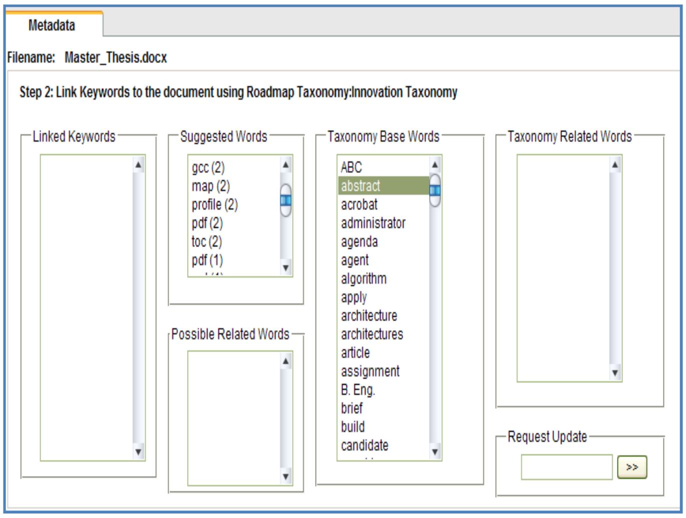


Figure 22: adding keywords with EDEN

- Summarization and comments (paragraphs)
  - Some tools provide also possibilities for users to add summary of their documents (e.g. in the metadata system EDEN™)

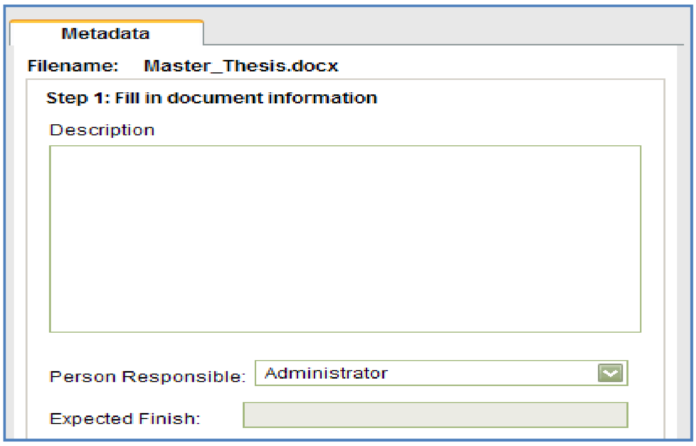


Figure 23: adding a summary describing the content of a document with EDEN

  - The actions of posting a comment become usual on blogs or on online stores, as to react directly to an article or give his advice on a product.

- Ratings systems (scores)

Generally, websites like Amazon, Ebay, Youtube… propose to advise but also to rate products or media, thanks to Users Generated Content (UGC) systems. According a recent survey published by IPSOS in December 2006, 25% of European Internet users and 33% of French users trust in UGC and ratings of other users (Lemeur, 2007).

Figure 24: ratings system of Youtube

VII.2. Assessment

The figure 19 emphasizes not only the feedback (i.e. the corrective actions, the enrichment by users) but also the need for measuring errors (or in contrary the effectiveness of the results).

We are going to present shortly the concept of performance and its measurement tools, the indicators. For this purpose, we will base our study on (Chauve, 2007).

VII.2.1. Performance

According to (Senechal, 2004), performance is not at the level of the result of the action, neither of the action in itself, nor even on the level of the objective, but it rather resides in the compromise between efficiency, effectiveness pertinence, and effectivity.

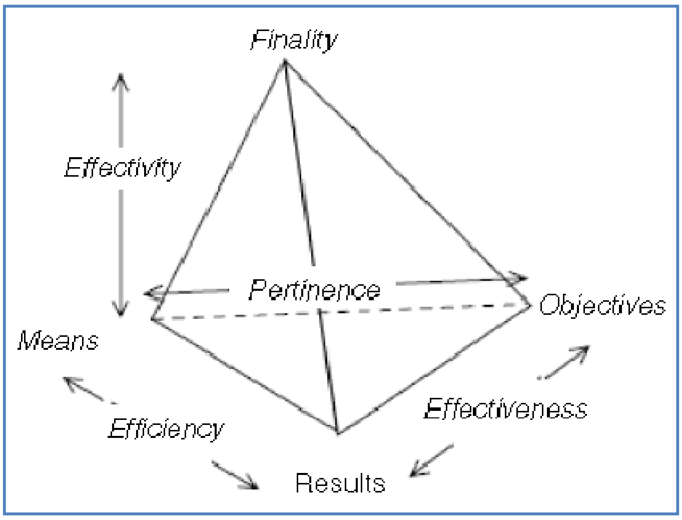


Figure 25: Tetrahedron of performance (Bescos, et al., 1995)

Among these four notions, two seem very relevant in the literature:

- The efficiency is the adequacy between the means and the results.
- The effectiveness is the adequacy between the results and the objectives.

The two last are generally too much abstract and thus hard to measure.

- The pertinence is therefore the adequacy between the means and the objectives.
- The effectivity is the adequacy between the objectives, the means and the results in comparison with the finality of the system.

Moreover, according to (Lebas, 1995), performance is not punctual but instantaneous in a dynamic flow. It is only a “picture” of the situation at a given time.

VII.2.2. Indicators of performance

According to (Lebas, 1995), again, the performance exists only if we can assess it, i.e. that we can describe it by series, or a vector, of measurements or indicators.

"An indicator of performance [or IP] *is a quantified data* [metrics] *which measures effectiveness and/or the efficiency of whole or part of a process or a system, compared to a standard, a plan or an objective, determined and accepted within the framework of a strategy of company"* according to (AFGI, 1992).

According to (Iribarne, 2006), a good system of measurement must be:

- Predictive: it reflects the performances to come as much as the last performances.
- Balanced: they cover all the fields of an organization or a system
- Communicating: they are used to communicate to the greatest number a vision future of the company and necessary evolutions.

Moreover, (Berrah, 2002) emphasizes two types or performance evaluation.

- Performance of instantaneous performance: you use the data which are available from the information system, survey, to acknowledge your capacity and choose.
  →In some extent it is the measure of effectiveness: do our systems or our organizations provide reliable results in comparison with our objectives?
- Progress piloting: the organizations want to have a continuous measurement in order to pilot their strategy.
  →In a way, that measures the efficiency: is our means or tools adapted for providing good results and how can we improve them?

To illustrate the use of performance indicators, we will give an interesting example focused on knowledge creation effectiveness (Moor, et al., 2002).

CATEGORY	CSF	INDICATOR	VALUE
Sympathized knowledge	Socialization	Direct communication links	100%
		Non-assigned working time	68%
		Regulated socialization	2,4%
Conceptual knowledge	Externalization	Number of bytes of project docs	47,5 Mb
		Percentage of hours assigned to project meetings	15%
Systemic knowledge	Combination	Number of categories in KB	3
		Number of items in KB	2071
Operational knowledge	Internalization	Number of years experience	9,6
		Frequency of use of KB	39,4

Table 7: examples of IP for knowledge creation process (Moor, et al., 2002)

VII.2.3. Assessment in the context of reification and virtual CoPs

We have observed that virtual CoPs have two fundamental needs: a need for information, in order to learn and innovate, and a need for people, in order to discuss and collaborate. Moreover, in our interpretation of LPP, we emphasized the needs for controlling automated reification and users' participation.

In addition, we can give the generic Information System (IS) success model of (Delone, et al., 1992), which emphasizes in some extent, that information quality and systems[22] quality impact on the quality of use and users (their knowledge, their participation and their satisfaction), then on the benefits of individuals and organizations.

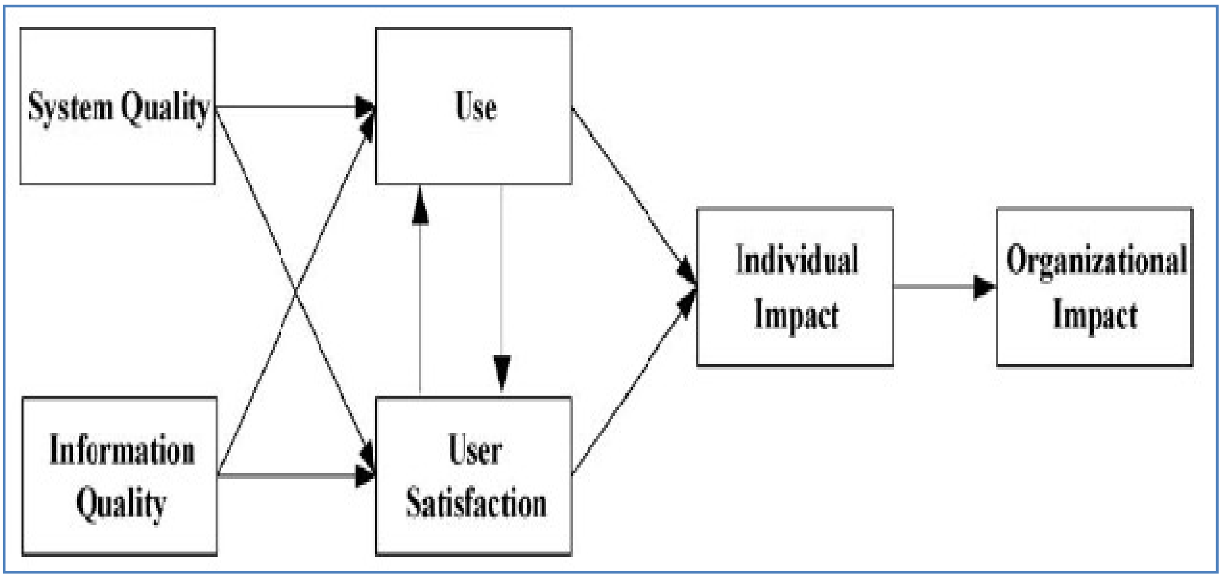


Figure 26: DeLone and McLean's IS success model (Delone, et al., 1992).

Subsequently, we can identify whose performance must be evaluated in the reification process, divided in:

- Static performance (information coming from corpora):
  - The quality of unstructured information or *"knowledge"* held in corpora must be assessed. According (PETERSEN, et al., 2005) ''Information quality'' is defined in terms of five characteristics of the information being exchanged: current, accurate, complete, consistently defined, easy to access
  - The status and the skills of users must be evaluated, with information held in corpora, in order to help to create collaboration and authorize participation.

[22] The system of reification carrying this information

- Dynamic performance (information added by automatic or human process):
  - Structured and modelled information map, must be assessed, i.e. we must check if reification process and systems are accurate and reliable or not. According (ISO/IEC9126, 2001) "System quality" can be defined by 6 characteristics decomposed in 29 sub characteristics.

Characteristic	Subcharacteristics
Functionality	Suitability, accuracy, interoperability, security, functionality compliance*
Reliability	Maturity,* fault tolerance,* recoverability,* reliability compliance*
Usability	Understandability, learnability, operability, attractiveness, usability compliance*
Efficiency	Time behavior, resource utilization, efficiency compliance*
Maintainability	Analyzability, changeability, stability, testability, maintainability compliance*
Portability	Adaptability, installability, replaceability, coexistence, portability compliance*

Table 8: Criteria for system quality

  - The assessment of user's participation updates the user's status' evaluation, in order to authorize user in further actions.

VII.3. Enrichment and evaluation, two complementary tasks

Quality assessment and enrichment are prominent to insure a reliable system that virtual CoPs can use. These two phases are narrowly linked:

- Some enrichment tools (like ratings) are used also to measure and assess, and the result of ratings can be then displayed to inform other users and enrich knowledge base
- As show in the figure 19, they are both necessary for control: user measures margins and errors (assessment), then he corrects in a feedback action (enrichment),
- they are both also important for explanation: enrichment and assessment can be used to recommend some point in the map or in the knowledge base and account for choices of system for extraction and visualization of some results.

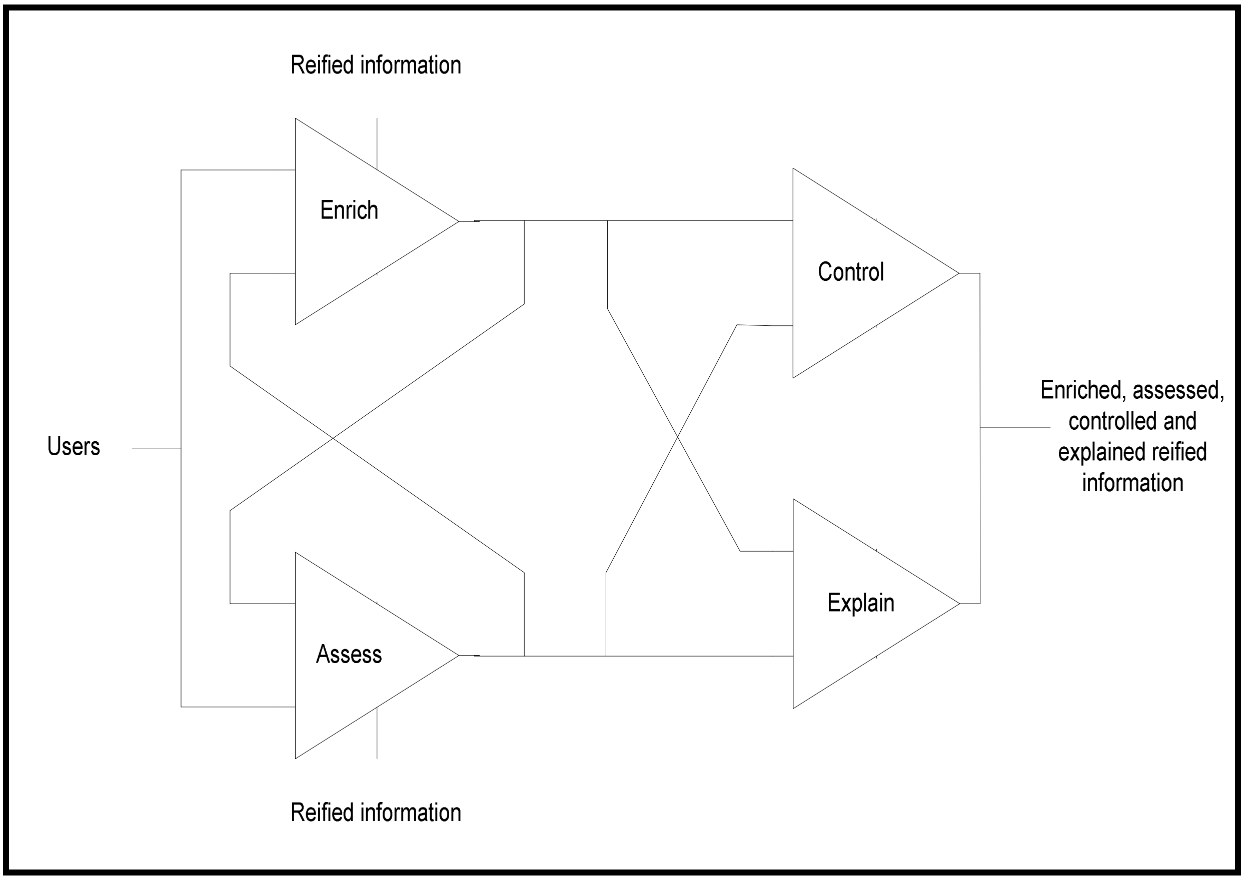


Figure 27: Link between manual enrichment and assessment

On the figure above, we can see the complementariness of enrichment and assessment, and their common roles for control and explanation.

VII.4. Conclusion

In this chapter we have emphasized the different possibilities for a user to control, explain and add value to reified information, and identified some criteria for the evaluation of information quality, user quality, system quality and quality of participation.

We have also seen the narrow link and the complementariness existing between enrichment and assessment. Thus, after the first step of automatic reification (where content analysis and visualization tools extracted, organized and displayed information about documents and people), we could consider human enrichment and assessment like a second step, in order to control and explain the automatic process.

This second step could be completed in a third step. Indeed, it would be interesting to know if we could automate partially explanation of reification process, and if the information coming from human participation could be automatically analysed to provide further information.

We are now going to present information filtering methods, which give some automatic tools to analyse user participation and help processes of enrichment and assessment.

CHAPTER VIII. INFORMATION FILTERING

Keyword: Recommender system, collaborative filtering, profile

We will present information filtering and its ability for enrichment and assessment.

VIII.1. Recommender systems

VIII.1.1. Definition

Recommender systems are information filtering (IF) tools which *"provide advice to users about items they might wish to purchase or examine. Recommendations made by such systems can help users navigate through large information spaces of product descriptions, news articles or other items."* (Burke, 2000)

According Wikipedia, a user's profile is created then it is compared to some reference characteristics, coming from the information item (the content-based approach) or the user's social environment (the collaborative filtering approach).

Therefore recommender systems are based on three keystones:

- The creation of a personal and customized user's profile
- The extraction of information about an item by content analysis (see section 1),
- The collaborative recommendation, where information is added by other users (for instance by rating or commenting) to help and ease choices.

The first element determines users' tastes, which are compared to the second point, and the last one takes in account the community's preferences.

Recommender systems use often user's ratings and average of ratings of a set of users to make some suggestions and sort the results.

So recommender systems are a means to assess and enrich content, in giving new information (under a form of advices) and an evaluation (with ratings…)

VIII.1.2. Two approaches, two benchmarks

From the previous definition, we can distinguish two pairs of characteristics:

Two approaches: Activeness and passiveness

User's profile or collaborative recommendation can be created by an active or a passive (automatic) way. Indeed, a user is able to fill himself his profile or make some suggestion, but his action give also some interesting information, although the user is not always aware of it.

Two benchmarks: internal source (history) / external source (neighbourhood)

Like in benchmarking process where the best practices are looked for in the history of the company (internal) or in competitors (external), the sources used by recommender system are twofold. They can come from the history of user's action as well as from those of other users. Therefore, the systems provide advice in taking account of user's habits and the tastes of the others.

VIII.2. Information filtering methods

In the literature, many different definitions and approaches are given for recommenders systems and collaborative filtering, which often mix the activities of a single user and the collaborative tasks. In our study, we decided rather to divide these notions in "individual filtering" (concerning the constitution of the personal user profile through the different actions of user presented in the previous chapter) and collaborative filtering (in focusing more on its ability for communicating and sharing information, and gathering people in groups of interests).

VIII.2.2. "Individual filtering" and "user profiling"

Passive filtering

A method that is thought to have great potential in the future is passive filtering, which collects information implicitly.

For instance, a web browser can be used to record a user's preferences by following and measuring their actions (Goecks, et al., 2000). These implicit filters are then used to determine what else the user will like and recommend potential items of interest.

Implicit filtering relies on the actions of users to determine a value rating for specific content, such as:

- Purchasing an item
- Repeatedly using, saving, printing an item
- Refer or link to a site
- Number of times queried

The passive filtering can also use information coming from content analysis (Ryszard Kruk, et al., 2005), such as:

- mailing-lists posts,
- links on home pages,
- citations in publications,
- co-authors of articles are utilized.

An important feature of passive filtering is using the time aspect to determine whether a user is scanning a document or fully reading the material. The greatest strength of the system is that it takes away certain variables from the analysis that would normally be present in active filtering. For example, only certain types of people will take the time to rate a site, in passive filtering anyone accessing the site has automatically given data.

Active filtering

In active filtering, the user takes active part in creating a set of his preferences.

He can give information about himself interacting actively with the system[23], for instance in giving directly his tastes to the system, in rating or commenting some items. This filtering is said to be active because the user is aware that his actions help the system to know about his preferences and to filter future items.

User profiling

This individual filtering leads to the creation of a user's profile, based on the user's preferences, collected passively or actively.

The nature profile and the importance of time aspect in passive filtering emphasize the "internal" and the historical dimension of the sources, i.e. the history of the user's actions. This explains the dynamic characteristics of these profiles.

[23] cf. Part VII.1.2

VII.2.3. Social or collaborative filtering

Definition

Within information science and human-computer interaction (HCI) research, a paradigm for categorizing, filtering, and automatically recommending information has emerged, called *"collaborative information filtering"* (Malone, et al., 1987). This approach is based on collecting and propagating word-of-mouth opinions and recommendations from trusted sources.

(Goldberg, et al., 1992) give the following definition:

"Collaborative filtering [CF] *describes all techniques leveraging incomplete information about tastes and opinions of a set of users".*

(Herlocker, et al., 2000) insist on the collaborative and predictive characteristics of CF, emphasizing its ability to *"predict a person's affinity for items or information by connecting that person's recorded interests* [user profile] *with the recorded interests of a community of people and sharing ratings between likeminded persons".*

Therefore collaborative filtering uses and gathers the information coming from the individual filtering, to give further information, more collaborative (like average ratings for instance). The "power of majority" and the sharing of views are prominent in this technique. Moreover, this method can also group people in comparing the similarity of their interests.

The most popular major types of the collaborative filtering are Active and Passive Collaborative Filtering. The distinction is based on the activeness of the user that receives information based on collaborative filtering (Ryszard Kruk, et al., 2005).

Passive collaborative filtering

The passive collaborative filtering is the aggregation of information coming from the actions of the many readers who access the system.

It is called "in-place" or "passive" because there is no direct connection between a person, casting a vote for instance, and the readers who come later and filter documents based on this aggregated information (Maltz, et al., 1995).

These "passive" indicators could be for example the number of visits of a website, or the average rating for an item…

Obviously these indicators are limited by the number of participant: to give an accurate rating and prevent from some error, a critical mass of users is required.

Active collaborative filtering

Another approach to collaborative filtering, builds on the common practice where people tell their friends or colleagues of interesting documents.

It is called "active" collaborative filtering because there is an intention on the part of the person who finds and evaluates a document to share that knowledge with particular people (Maltz, et al., 1995).

Unlike passive collaborative filtering, the benefits of "active" collaborative filtering are not based on a critical mass of users but on the diversity of their actions.

VII.2.4. Example

To illustrate and understand information filtering, we propose to analyse the possibilities and the information displayed on a well know website, YouTube.

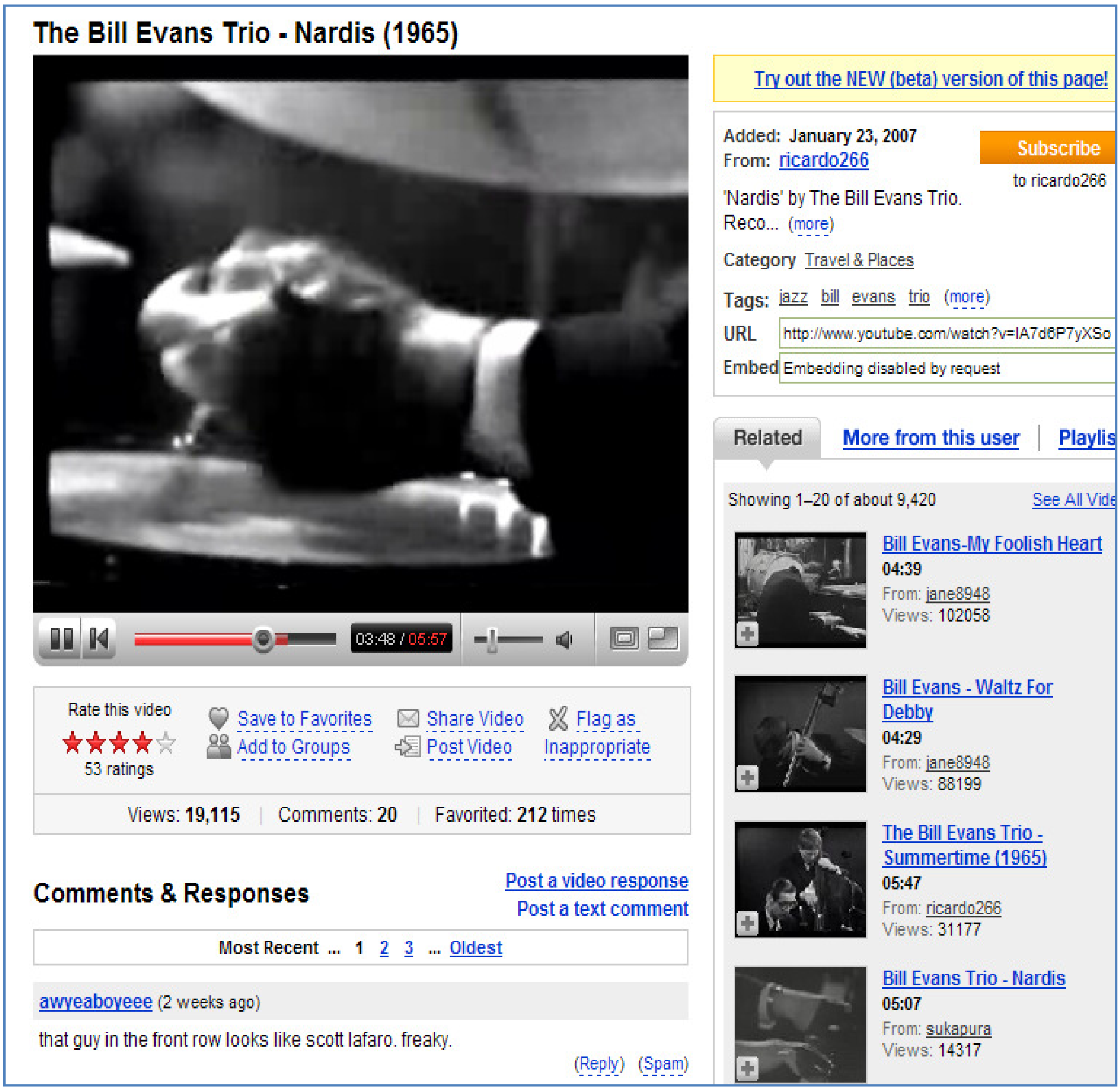


Figure 28: Example of information filtering on YouTube

- *Text comments* and *video responses* are Active Filtering (user is active) and Active Collaborative Filtering (user intentionally wants to share his point of view)
- The *average rating* (aggregation of votes), *the related media* (linked to the current video by tags), the *number of ratings*, the *number of comments* and the *number of bookmarkings* ("favorited") come from Active Filtering and are Passive Collaborative Filtering information (because the user is active but he is not aware or does not control the sharing of his actions)
- The *number of views* Passive Filtering and Passive Collaborative Filtering (because the action of viewing is not intentional for filtering and sharing an opinion)
- Finally, the possibilities for *adding the video to groups,* learning *more from the user who post the video* or *opening the profile of users who comments* ("awyeaboyeee") provide other information which contributes to the sharing of practices.

VIII.3. Information filtering for enrichment and evaluation

Information filtering can be used for enrichment and assessment:

By acting on corpora (passive or active "individual filtering"), user gives some information which characterizes texts:

- That enriches knowledge base (addition of information)
- active "individual filtering" (comments or ratings for instance) assess corpora

This information, symbolising the interaction of user with corpora, provides a users' profiles which can constitute dynamic customized indicators.

- This profile can be *"read"* by computer. Thus it constitutes a memory from where Artificial Intelligence can help user in suggesting likely or unlikely documents, and also evaluating the appreciation of a user on a specific content.
- This profile could be read by other users, to learn about people's interests, and assess collaborators.

These actions have also collaborative and explicative roles:

- According (Herlocker, et al., 2000), it can explain extraction of content analysis process (in justifying or not the display of information on the map coming from the corpus analysed)
- These actions help other users to find interesting documents (in adding information not emphasized by automation).
- Users can be gathered in groups of similar interests, that it ease collaboration

VIII.4. Conclusion

If information extraction automates classification of documents according their topics, information filtering proposes another classification: it sorts texts according the interests of user. This method is based on the memory of actions or data concerning the knowledge users. Computers learn about people, update profiles and use history of user's participation or tastes of other users to assess and enrich information.

Moreover, whereas participation constituted a manual enrichment and a manual assessment, information filtering can record and analyse this participation in order to assess then enrich automatically users and their activities, documents, reification.

Thus it gives obviously some answers to solve problems of LPP in virtual CoPs, especially at the levels of enrichment and assessment. Of course passive and active filtering should be well balanced, and another study should be led on the reliability of passive and active tasks.

To conclude, we will summarize the possibilities of IF in the following table:

	Individual filtering	**Collaborative filtering**
Inputs	Internal information, history	External sources, neighbourhood
Outputs	User profile, personal information	Collaborative information, "majority" opinions, similarity between different user profiles
Passiveness	Clicks, exploration, navigation,… (passive user)	Average ratings, numbers of clicks,… (unaware user)
Activeness	Comments, ratings, editing,… (active user)	Shared comments,… (aware user)
Assessment	• user (tastes, participation, habits) • documents (comments, ratings)	• reification (average ratings) • Evaluation of similarity of profiles
Enrichment	• Enrichment of knowledge base by user, • Computer learning about user (automatic analysis of participation) and automatic enrichment (addition of this information in the map for instance)	• Automatic enrichment of KB in explaining reification with collaborative information • Improvement of cognition (addition of new information, new link, creation of logic in the map for instance)

Table 9: Possibilities of Information Filtering

A survey on the main tools of collaborative filtering will be found in appendix 7.

We have approached the different steps allowing the automation, the enrichment and the assessment of reification. So, in the next part, this state of art will enable to propose a methodology and choose the most reliable tools supporting it.

Part C: Development

Methodology and case study

In this part, we are going to develop a methodology to solve the thesis problems emphasized in the first part, i.e. managing the reification process in a virtual Community of Practice in behaving Legitimate Peripheral Participation Process.

To do this, we will use our previous state of art, and we will study how we can combine the different techniques, from computerized reification to human interactions and automatic analysis of human interactions.

We will also focus on the different tools or methods explained before in the frame of our problem and propose a choice of the most efficient components in our “toolbox” to support our methodology.

Moreover we will suggest some indicators to assess information quality, user quality, process quality and user participation quality.

Then we will study how the software from Indutech can support this methodology, and we will finish in making a case study on a specific European virtual community of practice, VRL-KCIP. We will try out our methodology and the toolbox, theoretically in emphasizing what our propositions bring.

CHAPTER IX. PROPOSED METHODOLOGY

Keywords: Goals definition, Specifications, methodology, models, choice of tools, indicators of performance, Tree map

The problem will be approached in defining the goals to be reached, and constituting a kind of "book of specifications". Then we will be able to propose a methodology, which will be illustrated by some models so as to better explain it. We will finish by justify the choices for some tools supporting this methodology.

IX.1. Specifications

IX.1.1. Definitions of Goals

We have defined previously the aim of the study as the application of reification process in a virtual Community of Practice and follow Legitimate Peripheral Participation process. We have then divided this problem in three sub-problems, according our own interpretation of LPP, so as to reduce its complexity. In order to solve them, we will translate them in goals we will be able to specify and measure.

LPP interpretation	Thesis Problems	Goals/Needs
Peripheral Participation → Progressive learning	**Problem1.** How can we automate reification to ease CoP's progressive learning about information and people?	**Goal1.** Facilitating knowledge users to access to & learn about info and people, around an unstructured information in Knowledge Base
Legitimate Participation →authorized interactions	**Problem2.** How can we enrich and control documents knowledge base and reification process, when legitimate?	**Goal2.** Enabling human interactions and assessment in order to control et explain automatic reified information
Legitimate Periphery →expertise acknowledgement	**Problem3.** How can we assess and authorize users and their actions during the reification process?	**Goal3.** Assess users and their participation in order to authorize them to participate

Table 10: From the LPP to the definition of goals

IX.1.2. “Book of specifications”

Once the goals defined, we can specify them, to identify the different tasks we need to reach these objectives. To do this, we have used the previous state of art and our bibliographical report (Rauffet, 2007).

Thus, the first goal deals rather with:

- the framework design according to the methodologies observed in our bibliographical report (one must prepare and set the “Knowledge Base”)[24],
- the application of automatic reification, whose steps, content analysis[25], information architecture[26] and visualization[27], were shown in the first section of our state of art (to “facilitate access and learning by users”).

The goals 2 and 3 focus on the linked notions of enrichment and assessment studied in the second section, and their role for controlling and explaining documents and automated reified information. So it follows logically that the underlying tasks are related to users’ actions, Information Filtering and performance evaluation.

- The user can be helped for interacting with the system with the aid of dynamic tools[28], allowing navigation and active tasks, like creation, comments, ratings,…
- Effectiveness of given information and efficiency of automatic process[29] must be controlled, measured and communicated in order to insure reliability and give confidence to users
- Users themselves as well as their participation[30] must be evaluated, actively or passively, manually or automatically[31], in order to learn about them
- These evaluations are communicated and displayed in order to enrich the knowledge base and authorize users to interact with the system,

We have organized and detailed these observations in the following table, in identifying 6 specifications and 15 sub-specifications:

[24] Cf. (Rauffet, 2007), Ch.2
[25] Cf. Chapter IV
[26] Cf. Chapter V
[27] Cf. Chapter VI
[28] Cf. Chapter VII, VII.1.
[29] Cf. Chapter VII, VII.2.3.
[30] Cf. Chapter VII, VII.2.3.
[31] Cf. Chapter VII and Chapter VIII

Goals	Specifications
Goal1. Facilitating knowledge users to access to & learn about info and people, around an unstructured information in Knowledge Base	***S11. Organize the frame of Knowledge Base*** **S111.** Define the frame of KB, elements of corpora **S112.** Predefine users ***S12. Help users in automating process of reification*** **S121.** Limit the settings in extraction process **S122.** Automate organization of extracted information **S123.** Ease understanding with visualization tools
Goal2. Enabling human interactions and assessment in order to control et explain automatic reified information	***S21. Help users in interacting and adding information*** **S211.** Make the system dynamic to navigate among information **S212.** Enable users to create, modify, comment, rate, explain information from KB and automatic reification process ***S22. Help users in assessing information and process*** **S221.** Control and measure the effectiveness of information held in corpora **S222.** Control and measure the efficiency of the automatic process for extraction and display of structured and modelled information **S223.** Communicate and explain with the help of some performance indicators, so as to give confidence to users in the information and the system.
Goal3. Assess users and their participation in order to authorize them to participate	***S31. Assess users and their participation*** **S311.** Assess users, their use of the system, their tastes, manually or automatically **S312.** Analyse history of participation and similarity of interests ***S32. Use this evaluation*** **S321.** Communicate and explain with the help of some performance indicators, so as to give confidence to users **S322.** Enrich information about users and the potential links between them **S323.** Evaluate the degree of expertise of all participants and authorize them to interact with the system or not

Table 11: Book of specifications

IX.2. Analysis of specifications

We will now try to analyse deeper these specifications, in making a scenario-based analysis centred on the actions of users, in identifying precisely the different item emphasized previously.

IX.2.1. Scenario-based analysis

In order to deepen some details of the book of specifications and understand the possible activities of users, we have analysed the scenarios which would be probably used.

We have focused on the actions of users from the creation of artefact (manual reification of tacit knowledge into information), to the interactions with the documents and the maps (editing, browsing, comments, ratings,…).

We have identified 7 scenarios, as shown in the following figure:

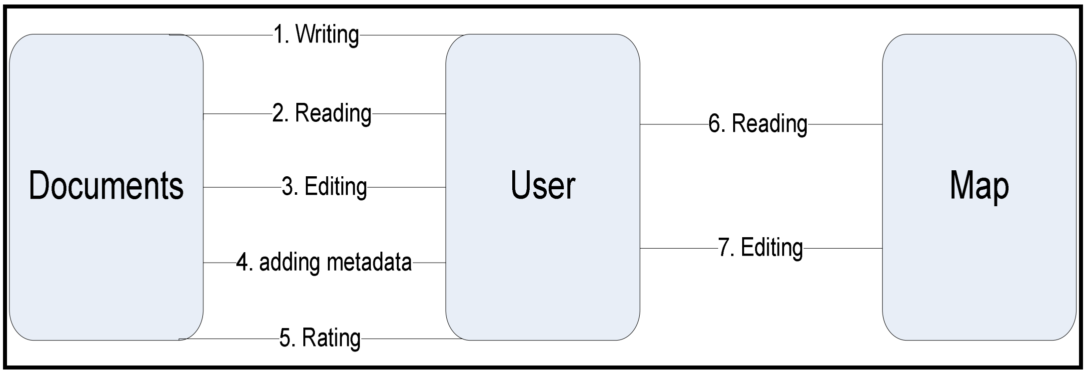


Figure 29: Scenarios of user's activities

The resulting scenarios are detailed as below:

1. User writing a document

- Translates his skills and his tacit knowledge into "explicit knowledge", into information
- Uses concepts
- Uses other documents
- Organizes his knowledge in artefacts

2. User reading a document

- Opens a document
- Looks for information
- Learns
- Controls and forms a subjective opinion about the content of the document based on its usefulness, accuracy, quality

3. User editing a document

- Adds information
- Corrects information held by document

4. User adding metadata to a document

- Adds some additional data about the content of the document
  - Adds keywords, tags
  - Adds paragraphs: he may post comments, summarizes the document
- Explains some concepts included in document
- Controls accuracy of information

5. User rating a document

- Puts a score to assess the value of the documents
  - Gives a value to a document
  - Sorts a group of documents according their range of relevance,…
- Gives an "indicator of confidence"

6. User reading map

- Browses map
- Looks for information
- Learns
- Controls accuracy of information

7. User editing map

- Adds information
- Corrects information displayed on map

IX.2.2. Identification of involved items

We will now attempt to identify the resources involved and emphasized by the "book of specifications". We have considered some items linked to documents and models (document, map, reification, group of documents), and some other "classes" related to user (user, participation, organization, group of users).

	Information	Actions
Document	"Hidden" concepts, authorship, date of publication…	***Active:*** comment, rate, modify, edited, modified… ***Passive:*** view, open… ***Automatic and collaborative:*** IF can aggregate this actions to give other indicators on documents
Map	Concepts, relationships, neighbourhood, dependence	***Active:*** edit, modify ***Passive:*** explore, browse ***Automatic and collaborative:*** IF can aggregate this actions to give other indicators on documents
Reification	Extracted and modelled concepts, explanation of extraction?	***Active:*** Previous actions can be used to evaluate efficiency of the process et effectiveness of processed information ***Automatic:*** Topic Modelling has statistics which could be showed to explain extraction?
Group of documents	Similar topics, similar authors…	Automatic reification completed by user's participation and computerized analysis of participation can gather documents into groups
User	***"static":*** name, location, organisation,… ***"dynamic":*** authorship, expertise, legitimacy degree, tastes, history of participation, group of similarity, participation type (passive, active)…	***Personal:*** User fills his own profile ***Collaborative:*** Other users give advices about the user (comments, ratings) ***Automatic:*** IF analyses participation of the user and the others to evaluate preferences, find similarity…
Participation/ Activity	Passive, active, individual, collaborative	IF analyses this participation
Organization	Members, location…	
Group of users	Similar tastes, common practices…	The different actions and IF can gather users into groups
Designer		Define KB, …

Table 12: identification of items emphasized by book of specifications

IX.3. Methodology and tools

First we will develop a methodology to apply reification, and automate, enrich and assess reified information in the frame of virtual CoPs. We will also build some models to understand its different steps. Then we will propose a choice of tools and techniques previously presented in the state of art, to support this methodology. Finally, we will compare our propositions with the book of specifications.

IX.3.1. Proposition of a methodology

In order to build our methodology, we will use:

- the description of reification process emphasized in knowledge Life Cycle adapted to Communities of Practices (cf. Fig.6), with three important sub-processes:

externalization
  - explanation (where knowledge is transformed into information with artefacts),
  - categorization (where unstructured information (free texts) is structured into data)

combination
  - personalization-contextualisation-modelling (where data is situated and linked to provide structured information)

- Some observations developed in the state of art
- our previous bibliographical report, especially for the study of KM methods, that the following table sums up (Ammar-Khodja, 2006; Rauffet, 2007):

Methods	Type of knowledge	Phases	Collection tools	Modelling tools	Results
MKSM	Domain knowledge	Framing, modelling, planning orientation	Interviews & study of documents	SADT, domain & task models	Knowledge book, orientation plan, KMS
REX	Activity Memory	Expert identification, Knowledge collection, extraction, modelling	Interviews & study of documents	Cartography and lexicon tools	Experience sheets, descriptive model & terminological model
CYGMA	Design Knowledge	Expert identification, Knowledge collection, textual modelling	Interviews & study of documents	Categorization of knowledge	Knowledge bases, professional reference frames
KADS	General / Domain knowledge	Expert identification, Modelling, Transfering to software programme	Interviews, observation & study of documents	UML	KBS
MOKA	General / Domain knowledge	Extraction, Modelling	Study of documents	Text, MML	Icare forms, formal models, KBS

Table 13: comparison of KM methodologies

- the scenario-based analysis, which outlines the different tasks or activity of the users
- the identification of involved items, which defines actors, inputs and outputs.

Step1: Manual reification

As pointed out on the knowledge life cycle, the first step is clearly the ***manual reification***, i.e. the scenario where users ***create artefacts*** in order to explain and transform their knowledge into information.

Step2: Framework design

The aggregation of the contributions of all users will constitute a corpus and will be the knowledge base with which people interact. This knowledge base must be ***configured*** by a designer in a step of frame design, like in many KM methodologies (for instance MKSM).

- **Framework analysis:** a designer should define the collaborative needs of the specific Community of Practice
- **KB configuration:** he should also choose the specificities of the knowledge base (format of documents,…)
- **Process and system configuration:** configure the settings of the different software. For example information, like list of authors or stop lists can be given to help and improve extraction.
- **User identification:** he could also gather information about users, in interviewing them (like in MKSM, REX, CYGMA and KADS methodology), and perhaps in creating some pre-profiles about them (names, emails, locations, speciality…)

Step3: Automated reification

Once constituted and configured, corpora must be ***analysed*** with computer, to ***extract*** relevant information inside them and ***ease cognition*** and retrieval by users.

As we have seen in the first section of the state of art, the automated reification is used for:

- **Content analysis[32]:** Analysing the content of documents in the knowledge base. Concept and relationships are extracted and categorized into data.
- **Information organisation[33]:** These concepts are then organised with the help of information architecture, which add some logics and create inference and so structured information.
- **Visualization[34]:** The structured information is finally displayed to ease navigation and cognition, and reduce complexity.

[32] Cf. ChapterIV
[33] Cf. ChapterV
[34] Cf. ChapterVI

Step4: Human-computer interactions

The automatic and structured information ought to be ***controlled, and explained***. To do this, ***users*** must ***enrich and assess*** information in interacting with KB and map, ***while authorized***. The second section of the state of art[35] and the scenario-based analysis point out the possibilities for this enrichment and this assessment:

- actively,
- passively,
- individually
- collaboratively,

with the aid of some tools supporting:

- dynamic navigation
- editing,
- comments,
- ratings,
- metadata feeding…

Step5: Analysis of human participation

To provide ***further information*** and ***assess participants***, human interactions can be analysed, with for instance user profiling and collaborative filtering.

This analysis can especially extract:

- Common views about information (average ratings…),
- user's preferences (dynamic and historic profiles following the history of participation…)
- similarity of tastes among several users (comparison and grouping of users according their profiles and their actions)

Step6: Explanation of automated reification

In parallel, automated of human-computer interactions and its automatic analysis, automatic reification could be explained and assessed, to give a better comprehension of the process. Indeed, ***text mining methods*** use generally ***statistic*** information[36], and it would be interesting to use some of them to understand why information was extracted and displayed.

[35] Cf. ChapterVII
[36] Cf. ChapterIV

Step7: Combined assessment

The assessments from user evaluation (Step4), automatic analysis of participation (Step5) and explanation of automated reification (Step6) can be ***combined*** and certainly ***weighted*** so as to provide good indicators for information about knowledge and people.

This assessment could be also used to determine which user can be ***authorized to participate***, following obviously the evolution of users to update the status of authorization.

Step8: Structured and combined enrichment

This combined assessment (Step7) and the manual and automatic enrichment (Steps 4 and 5) must be ***added to reified information***. Thus, information coming from corpora is finally completed by other information from users and their participation.

This new participative information could be added and ***displayed on the map*** for instance.

Loop

As a last observation, we can notice that this ***new information coming from user*** should also be ***controlled***. So it seems natural to put a loop joining the addition of structured and combined enrichment to the human-computer interactions.

We have summarized this methodology in the following table. We have also listed the actors involved by each step (human or computer), as well as the inputs and the outputs.

Finally, we have estimated the frequency of use of each step, differencing for example the initial and punctual step of frame design and the very usual tasks of users acting on the KB and on the map.

STEPS	ACTIONS	DESCRIPTION	ACTORS	INPUTS/OUTPUTS	Frequency
1	Manual reification	Transformation of tacit knowledge into artefacts, documents	User (human)	I: Tacit Knowledge O: Artefacts, information	Usual (user must be able to add document)
2	Framework design	Preliminary works, parameter gathering, system settings, design of inputs	Designer (human)	I : Needs of the specific CoP O: KB, pre-profile,…	Punctual (to initiate and adapt system)
3	Automated reification	Extraction, organisation, visualization	Computer	I: KB O: Structured information	periodic (to take into account the changes in KB)
4	Human Computer Interactions	read-only or creative, dependent of expertise degree and authorization	User (human)	I: Structured information, authorization O: Control, manual enrichment, manual assessment	Usual (to control and enrich)
5	Analysis of human participation	User profiling, Collaborative filtering	Computer	I: manual enrichment & assessment O: automatic enrichment & assessment	periodic (to take into account the changes due to participation and analysis)
6	Explanation of automated reification	Explain & understand computer's choices	Computer	I: results of content analysis and architecture O: automatic assessment	
7	Combined assessment	Use outputs of the steps 3,4 & 5 (computer created- and user generated-evaluation) and combine them to assess	Computer	I: automatic assessment O: information enrichment, user profile updates and user's authorization	
8	Structured and combined enrichment	Use outputs of the steps 4, 5 & 7 to provide structured and interactive information	Computer	I: automatic assessment, manual and automatic enrichment O: Enriched map, groups of common practice, groups of similar topics,…	

Control and enrich new information due to participation and automatic analysis

Table 14: Proposed Methodology

IX.3.2. Summary and models for the proposed methodology

In order to better understand this methodology, we realised a summary which explains what a researcher, the KM designer, the community of researchers and the system do. We drew also two models so as to illustrate it.

Summary: researcher's, KM designer's and system's points of view

Let us consider a ***researcher*** who writes a ***paper*** (Step 1, manual reification). There are already two kinds of sources of knowledge: the researcher who has tacit knowledge and the paper where the tries to express and formalize his knowledge (externalization). These two sources of information are the inputs of a ***system*** which aims at pushing information towards the other researchers in the CoP.

The ***KM designer*** constitutes an initial ***Knowledge Base*** with ***papers*** written by ***researchers***, and gathers information about these authors, in interviewing them and creating ***preliminary profiles***. He also sets different tools of the ***system*** (Step 2, Frame design).

The ***system*** really assists the ***community of researchers***, in extracting, organizing and visualizing on an ***interactive map*** the relevant information coming from the ***Knowledge Base***, constituted by the corpus of ***documents*** and the ***pre-profiles*** of the authors (Step 3, Automated reification).

This automation must be controlled, and the ***researchers*** must be able to ***interact*** with its results. Thus, a ***researcher*** opens documents, browses the map, comments, rates, adds and corrects the contents of different documents (Step 4, HCI). The question of ***authorizations*** between the researchers and the system further needs to be considered.

From this ***interaction***, the ***system*** can also provide some additional help for the ***researchers,*** in analysing the actions, completing the ***profiles*** of each ***researcher***, and determining the similarity of interests between researchers (Step 5, participation).

Moreover, the ***researchers*** can raise the question of the ***effectiveness of the results*** displayed by the automated reification. In order to give more trustable information, the ***system*** can use information coming from extraction, especially ***statistics*** concerning for instance the relevance of extracted information (Step 6, explanation of automated reification).

Finally, all the interactions of researchers, the analysis of participation and the explanation of automated reification could be combined, structured and weighted in a ***combined assessment*** of the knowledge base, the system and the researchers. This would result in several ***indicators of performance*** (Step 7, combined assessment). The assessment of a ***researcher*** is used for ***authorizing*** the researchers to interact in the CoP.

Moreover, all this ***new information*** from the participation of all the researchers ***enriches*** the content extracted by automated reification, in adding new data, or in giving trustable indicators on this extracted content with the aid of comments and ratings (Step 8, combined and structured enrichment).

SADT (Fig.29)

The first one follows SADT design. We aimed at show the links between reification (manual and automatic), human-computer interactions, and evaluation and enrichment made manually or automatically.

- **Reify:** Thus, tacit knowledge is reified, first by user (under artefact, document) then by computer (creation of a dynamic map)
  → Steps 2 and 3
- **Interact:** User interacts with system and KB. He learns (so he is "enriched") and he adds some further information onto KB and dynamic maps.
  → Step 4
- **Evaluate & Enrich:** This enrichment and the information coming from the reification in KB or dynamic map are evaluated manually or automatically (analysis of participation or explanation of reification) so as to:
  - Assess information held by documents in KB
  - Assess reification, i.e. information displayed by dynamic map (DM)
  - Assess users, following their progresses, then authorize them to act deeper
  - Assess their participation ("enriched KB and DM")

  → Steps 5, 6, 7
- The methodology gives:
  - Controlled and explained computerization (automatic reification process)
  - Enriched and assessed KB and DM
  - Assessed users
  - Authorizations for user's participation

Flow scheme (Fig.30)

In this model, we tried to show the different processes involved by our propositions. We build this flow based-model step by step in following the methodology. [37]

The final model represents the proposed methodology to solve the thesis problem. It emphasized the computerized reification, the human-computer interactions and the different activity, manual or automatic, for enriching and assessing reified information.

[37] This construction can be found in appendix 8.

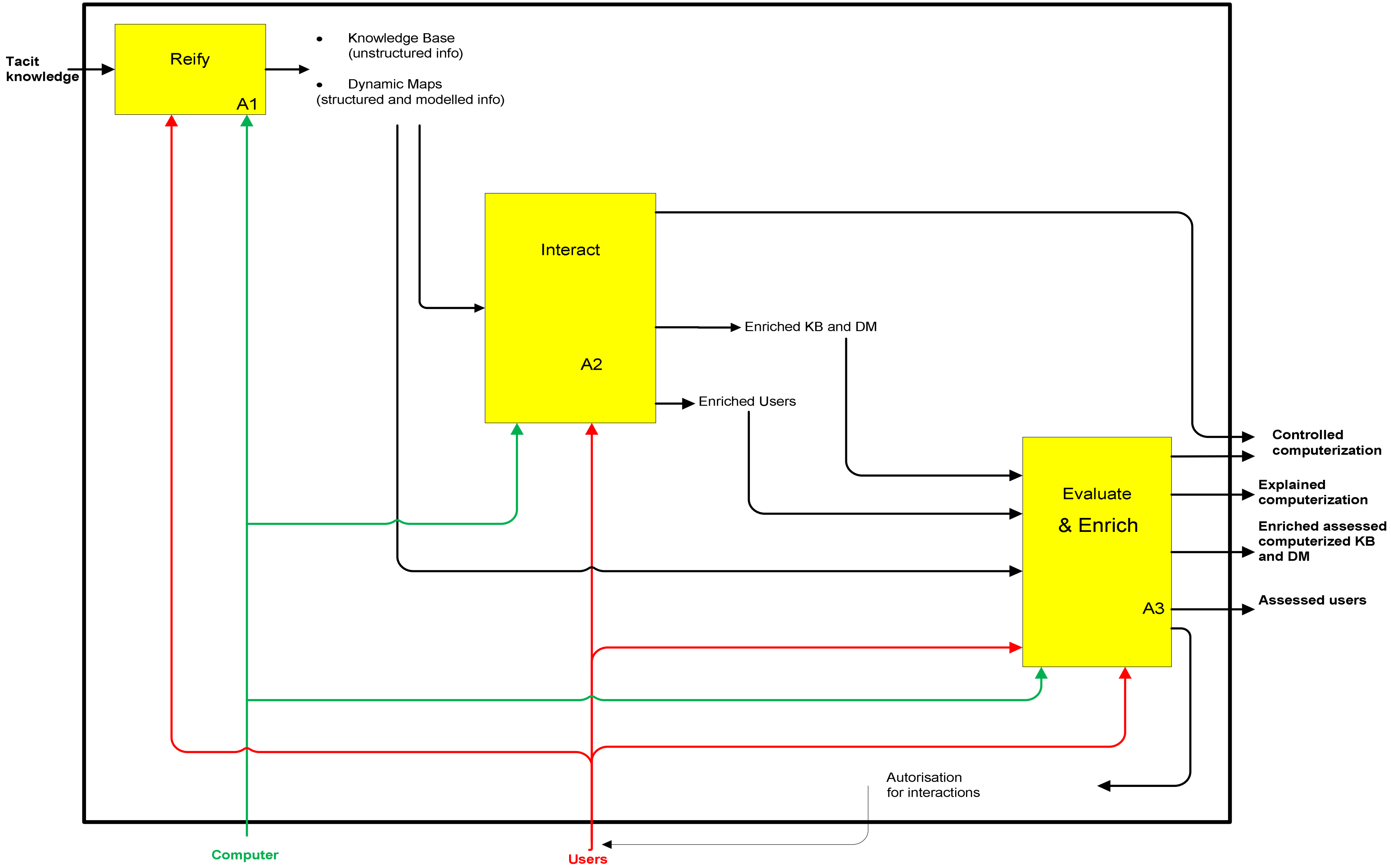


Figure 30: SADT, links between reification, interaction, evaluation and enrichment

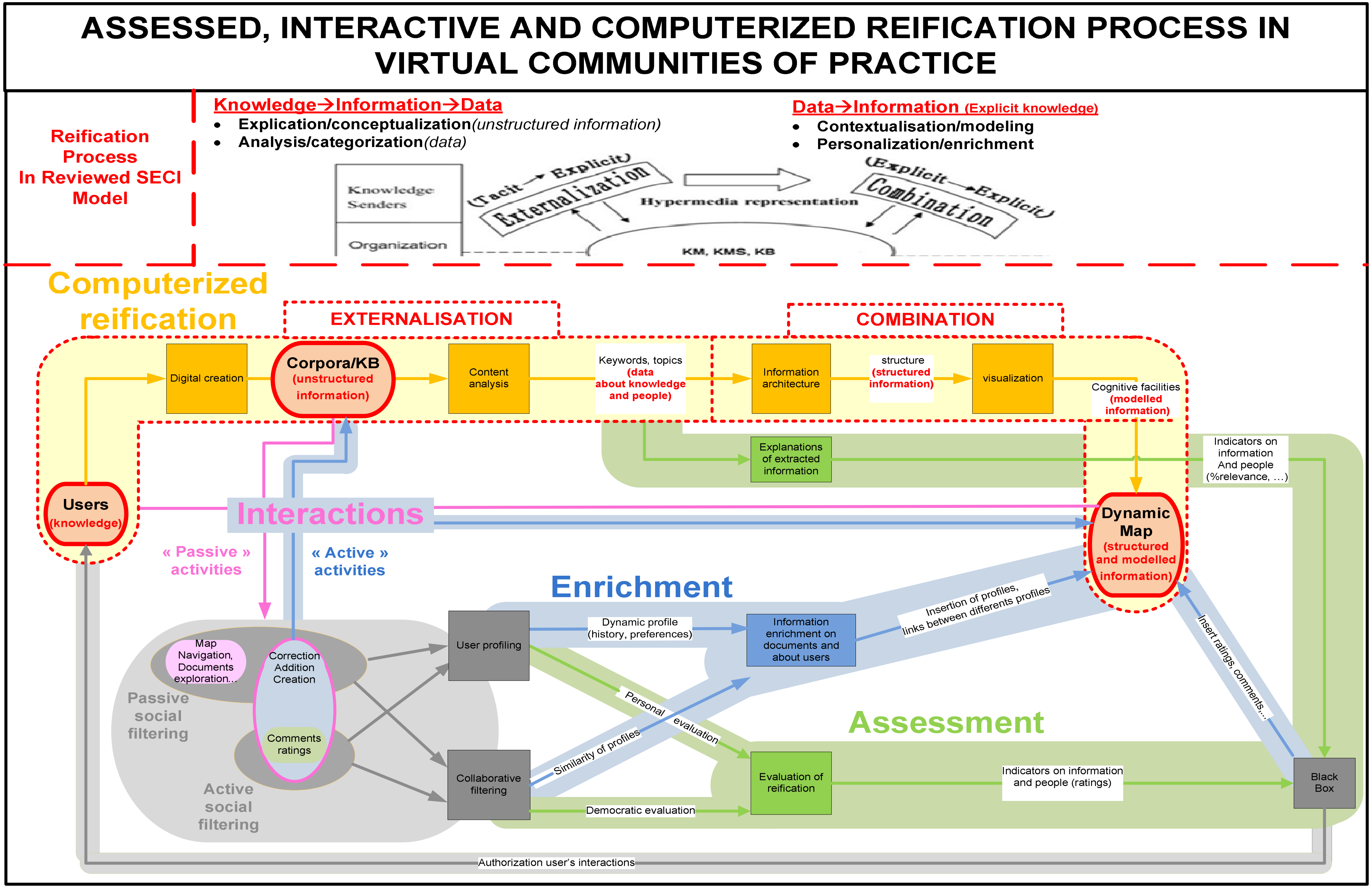


Figure 31: Flow based-model

IX.4. Tools & Methods

IX.4.1. Suggestions for tools and methods supporting the methodology

To support the proposed methodology, we need to find some accurate, efficient and reliable software. We will follow the steps of the methodology and use the state of art, in order to enlighten these “best” tools and methods.

Step 1, manual reification: We did not study manual reification tools in our state of art. Nevertheless, the tools used for digital creation are very well known, like text editors, table editors, scheme editors… (we can quote for instance Open Office or Microsoft Office, which provide a set of tools dedicated to digital editing).

Step 2, frame design: As emphasized in our bibliographical report, the step of framing and identifying user can be supported by interview, need analysis, book of specifications, and some language or model (for instance UML).

Step 3, automated reification

Extraction: Chapter IV emphasized text mining and topic modelling for content analysis and information extraction. We have seen that Topic modelling presents the advantages of skipping the lexical analysis and having unsupervised learning. Thus, the configuration (settings, modifications) is limited, unlike old text mining methods.

Organization: Chapter V outlines three information architectures:

- concept maps for the simplicity, the natural language and the user-friendly aspect,
- ontology for the addition of logic and the inference making,
- topic maps as the combination of benefits from concept map (natural language, semi-formal) and ontology (common and shared, high level for logic).

Thus topic maps are a good compromise providing both shared (ontology framework) and personal organization (oriented user interface), which can help CoP’s members.

Visualization: Chapter VI presents trees, maps and their combinations, and shows that cluster maps, fractal views and tree maps mix the advantages of trees and maps to ease navigation, reduce complexity and provide overview and detailed view. Overview will give the necessary background to user for learning, precise view could be adapted to the evolution of user’s expertise and personal needs. Moreover, tree map is a specific tool among that ones: it is a kind of multi-criteria map, which could be configured according user’s needs.

Step 4, human-computer interaction: To support this step, the tools for enriching and assessing are numerous, as quoted in Chapter VII.

- editing and exploration of documents can be done with the same tools of Step 1,
- editing and exploration of map use the possibilities of dynamics in cluster maps, fractal views and tree maps
- Comments, ratings, summarization … are possible with the help of metadata tools,…
- Collaborative filtering can support the sharing of personal point of views, in linking comments and ratings to media viewed by all users

Step 5, analysis of participation: Chapter VIII shows that the analysis of human interactions is possible with Information Filtering methods.

- User profiling gather personal and historical information about user and his actions, to determine his preferences and his habits
- Collaborative filtering can be used to aggregate individual actions (like ratings) and infer new information provided by all the users. He can also determine similarity and neighbourhood of taste among a community of user.

Step 6, explanation of automated reification: We have seen in Chapter IV that topic modelling gives some table of data and statistics in order to extract relevant information. These statistics from topic modelling could be perhaps used to explain better to users the “choices” of automated extraction.

Step 7, combined assessment: The different assessments (coming from manual ratings, Collaborative Filtering and Topic Modelling statistics) could be weighted in order to provide some relevant and reliable Indicators of Performance about effectiveness and efficiency of info, user, participation and automated reification

Step 8, structured and combined enrichment: Finally, all this new information, due to the enrichment and the assessment by users and computers, must feed the system and it must be organized onto map. We could also think about a scorecard displaying Indicators of Performance. The communication of this new information provides confidence to users about knowledge and people.

We put in the following table our propositions for the tools and methods supporting the methodology.

Methodology	Components	Advantages
Step 1	Text editor, Table editor, ...	Digital reification
Step 2	Interview, UML language, Need analysis...	Framing, user identification, configuration of the KB and the tools for automate reification and analyse of participation
Step 3	Extraction, organization, visualization tools	
	Topic modelling	- a few settings, - unsupervised learning text analysis - independent from lexical analysis
	Topic map	→shared (at the ontological meaning) and personal (non controlled vocabulary, user interface) →share common views on a subject (CoP LPP) →prepare visualization, help users with common practices and personalization
	- Cluster Map - Fractal view - Tree map (multi-criteria map)	- overview and navigation→ learning, discovering (for newcomers, or expert looking for area different from his domain) - precise/filtered view → expert recommended navigation, according experience and history - multi-criteria view → display according user needs (but difficulties to determine criteria)
Step 4	Human interaction (passive, active, individual, collaborative) Metadata tools, CF	- enrich - assess → control → explain
Step 5	CF, User profiling	- automatic aggregation - new information due to participation taken into account
Step 6	Statistic from topic modelling	→ explain and assess
Step 7	Performance Indicators	→assess users & docs, give allowances, control
Step 8	Data Management, Data mining, scorecard	→ combine and structure all new information to complete map

Table 15: Choices of tools and methods supporting the methodology

IX.4.2. Suggestion for a system of "indicators of performance"

Assessment is very tremendous in the methodology, to appraise information, users, automated reification and user's participation. However, the state of art does not bring about some clear propositions to support these evaluations. So it is why it could be interesting to try to suggest a system of Indicators of Performance.

Because the role of the methodology is providing information about concepts held in corpora and users, we will look for evaluating the effectiveness of information and users (comparison of results with expectations).

Reification and participation are the processes which bring this information and characterize users. So we will focus more on the efficiency evaluation of these two notions.

Performance	Indicators
Information (from corpora, maps, and user's participation) → **Effectiveness** (results evaluation)	- **Relevance** (individual and average ratings, % likelihood in topic modelling) - **Frequency** (number of views, number of comments, number of ratings,…)
Users → **effectiveness** (results evaluation)	**- Relevance** (authorship, acknowledgement by other people…) **- Experience** (number of publications, number of collaborations…) **- Status** (« trainee », « teacher », « reader », « actor ») - **Frequency** (attendance, contribution, activeness)
Systems and process of reification → **Reification's efficiency** (process evaluation)	**- Relevance of reified info** (comparison between explanation from content analysis and opinion from participants) - **Complexity** (number of Items in the KB, number of topics) - **Redundancy** (waste)
Participation → **Efficiency of user participation and authorization system** (process evaluation)	- **Relevance of added information** - **Frequency** - **Redundancy**

Table 16: System of Performance Indicators

IX.5. Conclusion

This methodology and the models emphasized two sources:

- ***the documents*** in Knowledge Base, where relevant information is extracted and structured with automated reification
- ***the users*** who interact with the KB and the map, in creating, editing, correcting… and whose the participation is also analysed, to provide further information.

Moreover, the proposed methodology and tools support Communities of Practice, especially in the Legitimate Peripheral Participation process, for overcoming barriers and limits due to digitalization and virtualization.

- ***Legitimate Peripheral Participation***
  - ***Peripheral Participation:*** Users can learn about and acquire common language and practice about people and concepts, in using externalized and combined knowledge. Indeed, automated reification and human-computer interactions ease the Peripheral Participation, the progressive learning, and respond to the problem due to the lack of face-to-face and situated discussions.
  - ***Legitimate Participation:*** When authorized, the interactions of users enrich, control and complete the automated reified information. That makes users confident in given, displayed information and the reliability of the system. This information will be used at a later stage in order to create trustable collaboration.
  - ***Legitimate Periphery:*** the user participation is analysed, to follow and measure user expertise. That provides information and assessment about users and authorizes them to participate in progressive actions, according to their expertises.
- ***Responses to barriers and limits***
  - ***Virtualization:*** The complexity due to the size of virtual CoPs and the unsynchronized environment is partially solved: people learn first with the system (that replaces the face-to-face communication and the situated learning), then they choose their collaborators to discuss further and innovate.
  - ***Digitalization:*** The automated reification overcome the problems for indexing and categorizes information and increasing corpora.

It is important to keep in mind the scope of the study, limited to the flow of reification. So all the collaborative and creative actions are excluded, because we have only focused on the way to structure, model and enrich existing information, but not on the way to create new knowledge and innovate.

We summed up these contributions in observing how the methodology (IX.3.) and the chosen components (IX.4.) fit the book of specifications (IX.1.).

Specifications	Methodology	Tools
S11. Organize the frame of Knowledge Base **S111.** Define the frame of KB, elements of corpora **S112.** Predefine users	**Step 2:** *Frame design*	KM methodology - Design - Interviews,…
S12. Help users in automating process of reification **S121.** Limit the settings in extraction process **S122.** Automate organization of information extraction **S123.** Ease understanding with visualization tools	**Step 3:** *Automated reification*	Reification - Topic Modelling - Topic Maps - Tree Map, Fractal Map, Cluster Map
S21. Help users in interacting and adding information **S211.** Make the system dynamic to navigate among information **S212.** Let users create, modify, comment, rate, explain information from KB and automatic reification process	**Step 4:** *Human-Computer interactions (HCI)*	Active & Passive interactions - Open documents, dynamic navigation of maps,.. - Editing, rating system, UGC, posts,…
S22. Help users in assessing information and process **S221.** Control and measure the effectiveness of information held in corpora **S222.** Control and measure the efficiency of the automatic process for extraction and display of structured and modelled information **S223.** Explain and communicate and with the help of some performance indicators, so as to give confidence to users	***Steps 5:** Analysis of participation* ***Step 6:** Explanation of automated reification* ***Steps 7, 8:** combined and structured assessment and enrichment*	Interactions & Performance evaluation - Rating, comments (individual & collaborative, active & passive) - Display of statistics information from topic modelling - understandable IP, Scorecard
S31. Assess users and their participation **S311.** Assess users, their use of the system, their tastes, manually or automatically **S312.** Analyse history of participation and similarity of interests	**Steps 4, 5:** *HCI and Analysis of participation*	Interactions & IF - Collaborative tasks, IF - User profiles, CF
S32. Use this evaluation **S321.** Communicate and explain with the help of some performance, indicators so as to give confidence to users **S322.** Enrich information about users and the potential links between them **S323.** Evaluate an expertise degree and authorize them to interact or not	**Steps 7,8:** *combined and structured assessment and enrichment*	Performance evaluation - Scorecard, understandable IP - Data management, data mining,… - IP, Scorecard, "black box" (for authorization)

Table 17: Methodology and tools versus book of specifications

CHAPTER X. THE METHODOLOGY AND INDUTECH'S TOOLS

Keywords: CAT, Organon, EDEN

In this chapter we propose to present Indutech's tools. Then we will study their potential assets to support the methodology. We will focus on three key tools of Indutech: CAT, a topic modelling tool, Organon, a conceptual framework browser, and EDEN, EDEN^{TM}, an Enterprise-wide Innovation Management Platform.

X.1. CAT, a topic modelling software

X.1.1. Presentation

Indutech's Corpus Analysis Toolkit (CAT) is a Human Language Technology (HLT) and Natural Language Processing (NLP) tool that aims to help users understand vast amounts of structured or unstructured text data. Features include topic/concept extraction, collocation extraction and using regular expressions to find almost any item you want (Authors, Collaborators, Institutes, etc.).

CAT works as follows:

- Firstly, electronic documents about the domain under study are collected and grouped in a corpus (with several sub-corpora if required).
- CAT is then configured to do some of the following analyses depending on the need:
  - Extract certain patterns (e.g. e-mail addresses, dates, website URLs, capitalized terms, etc.) from the corpora specified.
  - Extract significant one-word (milling) and two-word terms (milling machine) from the corpora specified and rank such terms using a significance score.
  - Arrange the documents in a number of topic (the number is determined by the user) based on the content of the documents contained in the corpora specified.
  - For each topic found, give a profile in terms of the one-, two- and three-word terms that describes the topic. Overlaps between topics may further be determined.

- Per topic found, give a probability that a given document corresponds to a given topic (this is done for all documents and all topics).
- Find the conceptual similarity between all the pairs of documents in the specified corpora.

* For very structured documents (e.g. academic papers, certain websites, etc.) CAT may be configured to extract the desired information and semi-automatically populate it in a CF as specified by the user.

In the near future Indutech team is looking to incorporate named entity extraction (people, places, etc.), concept-based searching and automatic inference of relations between named entities.

X.1.2. Tests and analysis of the efficiency and the effectiveness of CAT

To better emphasize the assets of CAT and its potential benefits for the proposed methodology, we will focus on the efficiency and the effectiveness of the tools, in order to study if it provides fast and relevant information.
To do that, we made and use the findings of some tests. The use of this software is very simple: you have only to constitute a corpus of texts, and rules the number of topics you want.
<u>Efficiency</u>
To measure the efficiency, we tested CAT on some different corpora, in varying the number of topics. We aimed at understanding the comportment of the software and showing if it is adapted to perform information extraction.

The measured variable is: Time (in seconds).

We also observed the influence of several parameters:

* S: the size of the corpus (number of words),
* N: the number of documents,
* L: the average size of a document (number of words)
* K: the number of topics.

The data of the tests will be found in Appendix8.

From these results, we can observe that:

- The time of processing increases with the size of the corpus (S=N*L), a number of topics (K) given. We can observe that the functions are almost linear.

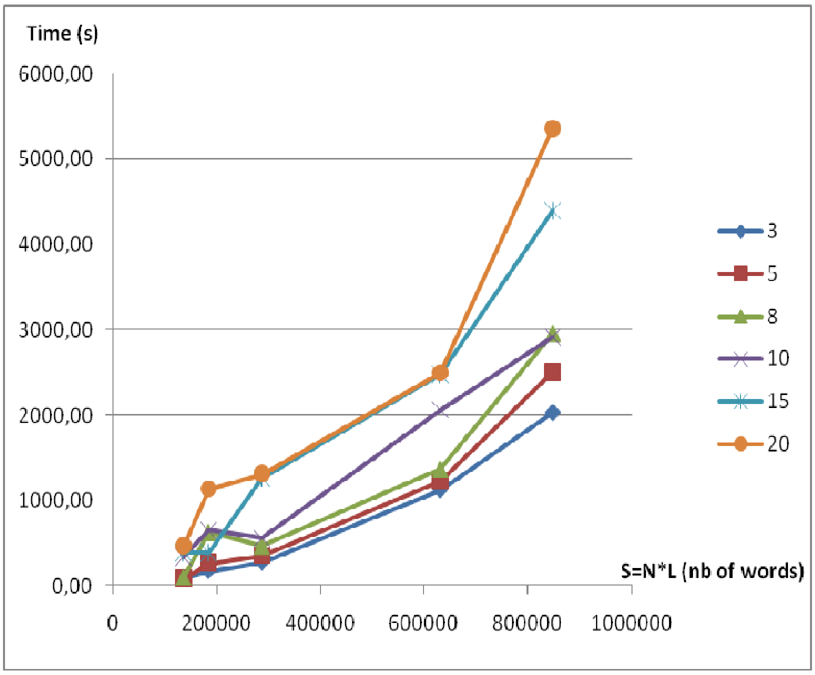


Figure 32: Influence of the size of corpus

- The time of processing increases with the number of topics (K), a corpus given.

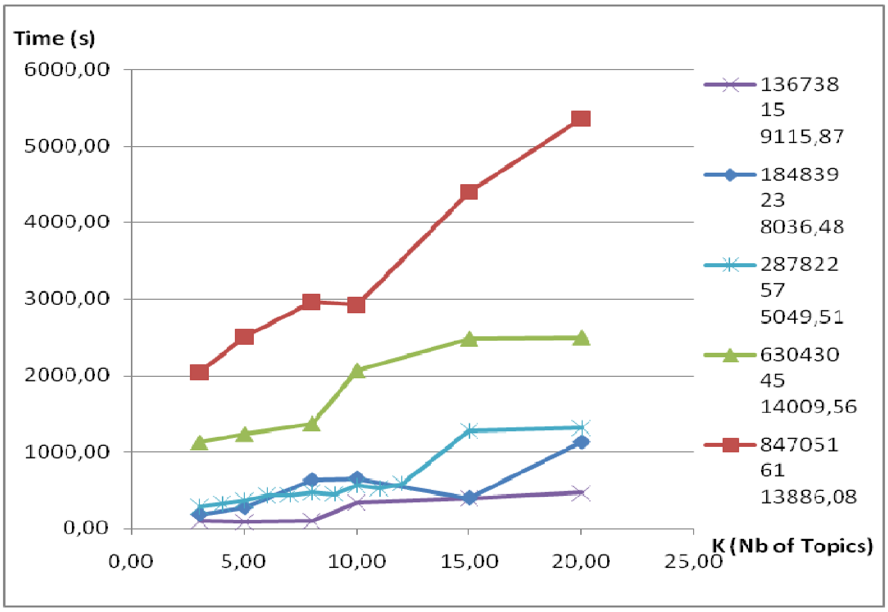


Figure 33: influence of the number of topics

- Taken alone, the parameters N (number of documents) and L (average length of a document) are not relevant. But combined (S=N*L), we have seen their influence on the time of processing. In the same way, we attempted to combine S and K to determine their linked influence.
  - The time of processing seems to increase with S*K, but the results are too dispersed. It means certainly that the parameters are not correctly weighted.

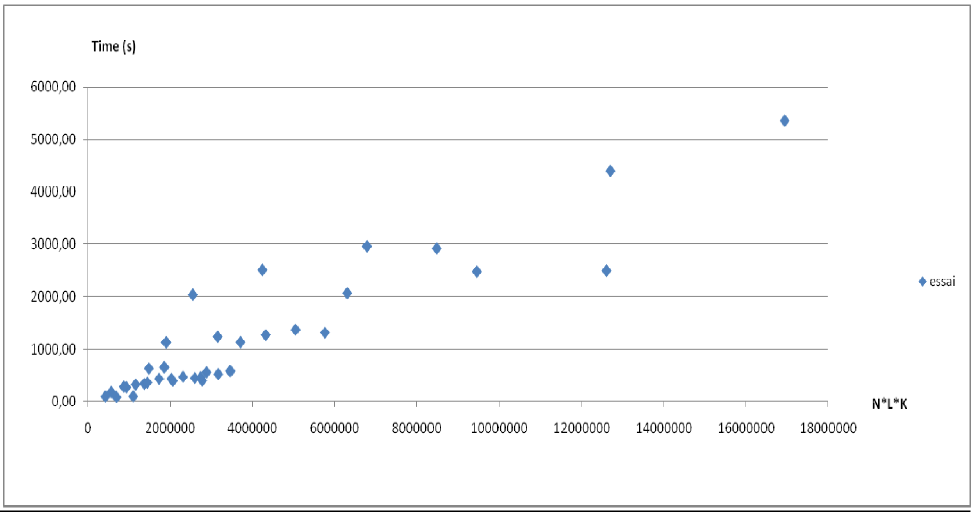


Figure 34: Combined influence of S and K

  - We tried several tests in varying the value of x and y in the formula K^x x S^y. (x,y)= (0.5, 1) seems to be better, having almost a linear comportment. In addition we can assume that S is prominent on K.

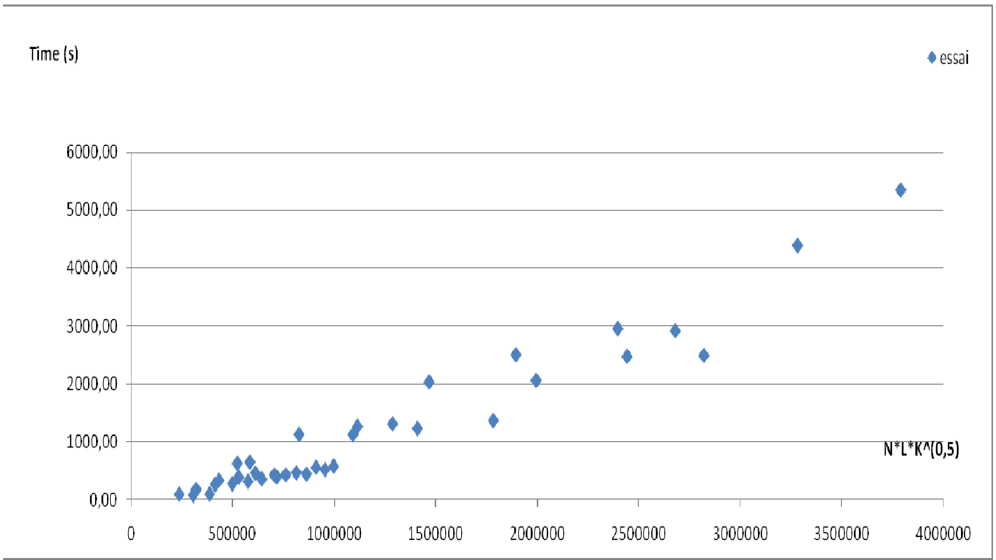


Figure 35: Time of processing ~N L $K^{0.5}$

This study is necessary, because it is important to understand the comportment of CAT so as to have a strategy to design the Knowledge Base and to configure Topic modelling, and so improve the efficiency of the automated reification.

Moreover, we can also notice that for little and average corpora (between 15 or 150 texts), the time of processing could reach big values (33105s for a corpus of 145 texts and 3 millions words).

That confirms the assumption that Knowledge Base must be well designed, and that the process of information extraction must be periodic but not continuous, because it can be long.

<u>Effectiveness</u>

To measure the effectiveness of CAT, we will present the results of a Case Study lead by Wilhelm Uys from Indutech (Uys, 2007), on a corpus of the knowledge network CIRP.

CIRP is composed by 500 members from 46 countries and it is organised in 10 Scientific Technical Committees (STCs) distributed by knowledge domains. The tested corpus is composed by 613 documents coming from the different STCS.

The Case Study shows that CAT is able to (cf. Appendix 9):

- **Automatically group papers into meaningful categories based on their content**: *Generally, the dynamic classification based on the concept was better than the CIRP classification based on the department belongings.*
- **Determine descriptive terms for each category of documents:** *The descriptive terms followed the classification of CIRP in STC, emphasizing thus the relevance of the results from CAT*
- **Determine overlaps between categories in terms of descriptive terms:** *The topic overlaps reveal that the shared words between topics are mostly words signifying abstract concepts, like 'system', 'process', 'cutting', and so forth. Also, the number of times a word occurs in a topic overlap gives an indication as to its prevalence.*
- **Determine which papers are conceptually similar to any given paper**: *The similarity results were reasonably accurate after comparing manually the texts.*
- **Determine descriptive terms for each paper:** *The descriptive terms per paper provides a very good characterisation of the content of the paper. Both ends of the spectrum are covered: the single-word key terms provide a more abstract, general view on the paper, whereas the two-word key terms give a more specific, detailed view on the paper, mostly consisting of the jargon of the subject at hand.*
- **Determine descriptive terms for each STC:** *The key terms provide an excellent characterization of the documents in the collection. The most highly ranked two-word key terms give an idea as to which specific techniques are mentioned most often in a given field of study.*

- **Automatically extract metadata from papers:** *The metadata extraction works well in general*
- **Determine descriptive terms for each author:** *the comparison of the content of the papers of an author and the author's assigned keywords agree very well*
- **Determine in which STC a given paper would fit best:** *Because the underlying categories extracted by CAT don't fit exactly the STCs' group of paper, the best representative paper fits more a given underlying category than a given STC.*
- **Determine outlier papers for a given STC:** *Results should be compared by an expert*
- **Determine the most conforming papers for a given STC:** *Results should be compared by an expert*

X.1.3. Assets for supporting the methodology

Obviously the CAT system from Indutech provides a **Topic Modelling** tool for the **extraction of concepts and relationships** in unstructured texts. Thus CAT supports especially the "**Information Extraction" in the Step 3** of the proposed methodology, the automated reification.

The time of process can be long, thus **it is necessary to design the Knowledge Base and set the software cleverly**, knowing the comportment of CAT in function of the length of corpora and the number of topics.

According the case study of Wilhelm Uys, The **results are quite accurate**, and CAT provides large possibilities for knowledge networks, in discovering hidden characteristics, relationships and similarity among papers from different domains.

X.2. Organon, a conceptual framework browser

X.2.1. Presentation

Organon was developed as a tool to help its users to structure unstructured, qualitative problems to gain an understanding of the problem without losing the context of the detailed elements of the entities in the makeup of the problem. The network structures that can be created, explored and edited using Organon are called conceptual frameworks.

Using Organon, various users can explore the conceptual framework (CF) of a certain environment, starting from an entity known to them, and reaching related, unknown entities by clicking on self-explanatory relations gaining an understanding of the environment as they go along. Users can further expand the CF by adding new entities or by associating existing entities - using new or existing relations - adding their own understanding to the CF while doing so.

Organon has the following features:

- Explore the CF - by clicking on entities and associated relations - to gain an understanding of the environment represented.
- Expand the CF by relating existing entities to new entities using existing or new relations.
- Search for all entities matching a given string and start exploring the CF from any entity in the result set.
- Find how two sets of entities are connected in the CF.
- Build a tree view starting at a given entity and expanding on specified relations.
- Create templates for frequently used entities to speed up the process of expanding
- the CF.
- Get more information about a given entity in one click by using the Lookup on Google or Lookup on Wikipedia commands.

X.2.2. Assets for supporting the methodology

Organon provides two main assets:

- A better information architecture, allowing to structure unstructured information with the aid of a network infrastructure
- A good visualization, helping user cognition, with the help of possible detailed view, overview, and Conceptual Framework Browsing.

Seemingly, **Organon** can give some technical responses and support in some extent the **step 3 of automated reification**, especially for **information architecture** and **visualization.**

X.3. EDEN™, an Enterprise-wide Innovation Management Platform

X.3.1. Presentation

EDEN and its Internet version WEBEDEN are software supporting a multi-disciplinary team through a variety of wide change projects within a company. They enable the team to follow a pre-defined structure, which acts as a Roadmap through their particular project. Thus they provide the team with relevant, good practice information that is easily accessible and easy to share. They will also capture experience and knowledge gained by the team members, ensuring repeatability and a learning culture.

Actually, EDEN^{TM} and WEBEDEN^{TM} aid their users in:

- ***Document Management in Context:*** With the Roadmaps, they create a structure in which users can store the documentation and information of projects they work on. That includes:
  - file management (folders, files),
  - file security
  - document version control.
- ***Knowledge Management:*** In addition of their storage structures which ease information retrieval and capitalize new knowledge generated by a project, they provide some metadata tools and some search methods.
- ***Project and Programme Management:*** The use of the Roadmaps structure for collaboration help teams and Programme Managers to conduct projects in a structured way, with necessary background information, and assessment milestones.

X.3.2. Assets for supporting the methodology

Thus EDEN^{TM} and WEBEDEN^{TM} are collaborative platforms, which provide:

- A document architecture (based on roadmap) and a visualization (the hierarchical tree of roadmap)

→ That contributes to **step 3**, especially for **information architecture and visualization.** Indeed, the user's classification according a project structure can complete the hierarchy made out by the automated reification.

- A virtual platform allowing the interactions of users, especially for enriching content, adding metadata, correcting and updating documents…

→ That helps the implementation of the **steps 4 and 5**, in supporting the **Human-Computer Interactions and the recording of these actions** ("document version control")**.** Already, the user's classification (put files in folders and in specific step of a roadmap) can be seen as passive filtering.

- Finally, with their file security management, they could be a base for organizing the user's authorizations system of the methodology.

→ That provides a frame for implementing the **step 7**, for **user's authorizations** after assessment.

X.3. Conclusion

In this Chapter, we have attempted to demonstrate that the software product of Indutech could partially support the proposed methodology. So CAT, Organon and EDEN could be incorporated into the proposed toolbox, because they provide some advantages which fit with the needs expressed in the methodology.

Eventually, we attempted to figure out the partial combined use of these three presented tools in the frame of our methodology in the following scheme.

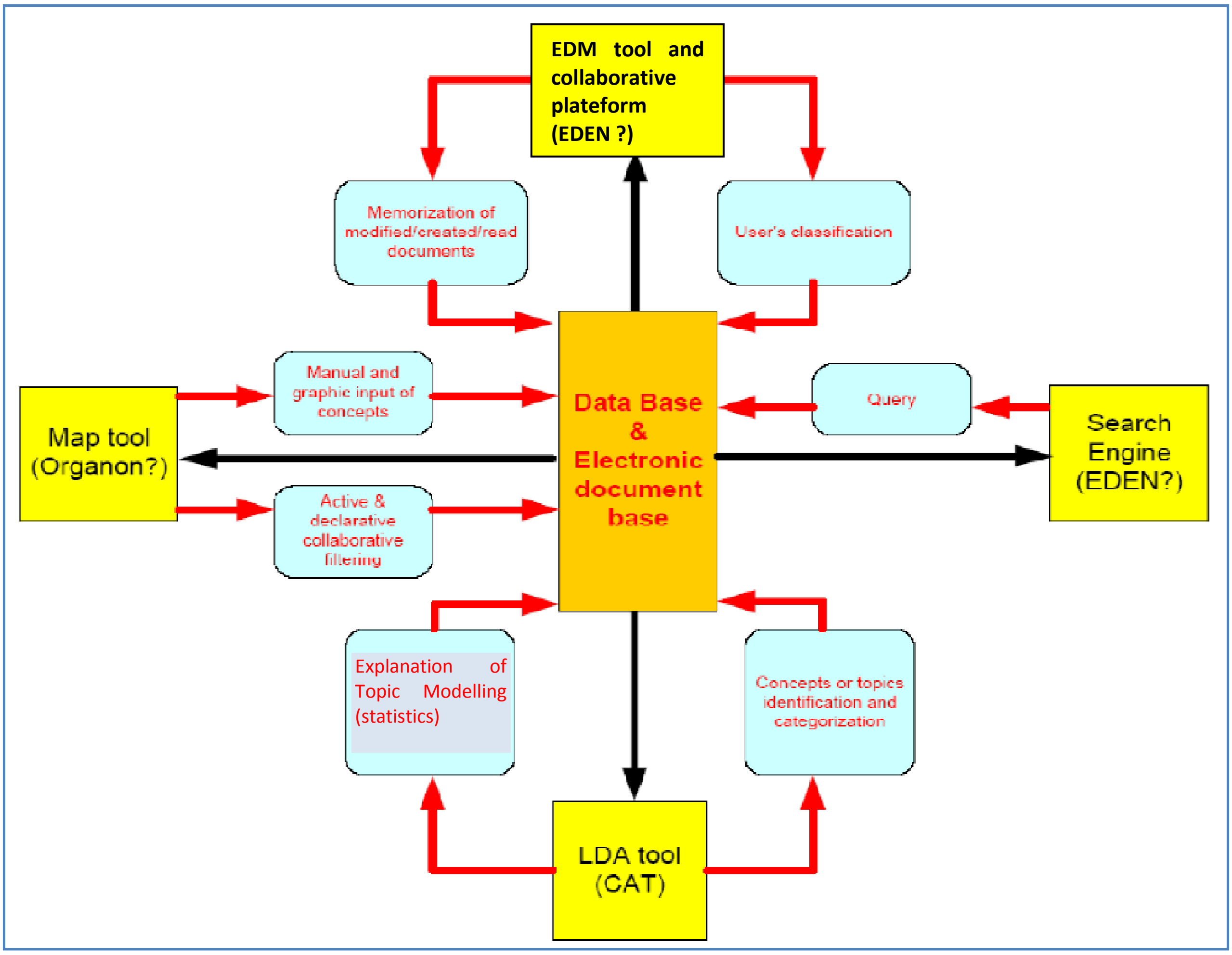


Figure 36: Combined use of CAT, Organon and EDEN

CAT, the topic modelling tool, extracts concepts and categorizes unstructured texts by topics. It can also provide some statistics to explain the process.

Organon, as interactive and dynamic map tool, allows for browsing and editing the visualization of the concepts and their relationships, with documents, authors, departments,…

EDEN^{TM} could be used both as a Electronic Management Tool (for organizing files, managing the security…), a collaborative platform (for supporting the users' interactions) and a Collaborative Filtering limited tool (for the management of document version control and in recording the different updates of a document by a user). Finally, the search methods of EDEN can be used to find the information needed by the user.

All the processes, from computer or human, feed the database which then distributes this new information into the different system.

To conclude, we provide as below a table summing up the assets of the different Indutech's tools for the proposed methodology:

Indutech's tools	Assets	Linked steps
CAT	- Topic Modelling tool for - extraction of concepts and relationships in unstructured texts - discover of hidden characteristics, relationships and similarity among papers from different domains - accurate results (effectiveness)	- Step 3 (Information Extraction)
Organon	- network infrastructure - possible detailed view, - possible overview - Conceptual Framework Browsing	- step 3 (information architecture and visualization)
EDEN^{TM}	- Document user's classification along a roadmap - Virtual platform, metadata system - Document version control - File security Management	- step 3 (information architecture and visualization) - steps 4 and 5 (Human-Computer Interactions and the recording of actions) - step 7 (user's authorizations)

Table 18: Assets of Indutech's tools for the methodology

CHAPTER XI. THEORETICAL CASE STUDY ON VRL-KCIP

Keywords: Case study, VRL-KCiP, Theoretical analysis, Tests, Topic Modelling, Fractal view, Tree Map,

To finish to put in context the methodology, we will present in this final chapter a specific virtual Community of Practice, the European VRL-KCiP Network of Excellence. We will observe theoretically how the proposed methodology can fit the needs of VRL-KCiP.

XI.1. VRL-KCiP, a virtual Community of Practice

According its website (VRL-KCiP, 2007) ,VRL-KCiP is a European Network of Excellence created in order to reduce the fragmentation of research in the field of production technologies. Its aim is to support dynamic organisations, inter-enterprise operability, and necessary standardisation.

Figure 37: VRL's Logo (VRL-KCiP, 2007)

To do this, it brings a multicultural approach to the integration of modelling and simulation of knowledge-based production processes on the one hand and to the relations between the joint partners on the other.

Indeed, it established a delocalized research structure in the area of Holistic Production Research at the European level which currently consists of 24 internationally famous research sites from 15 different countries.

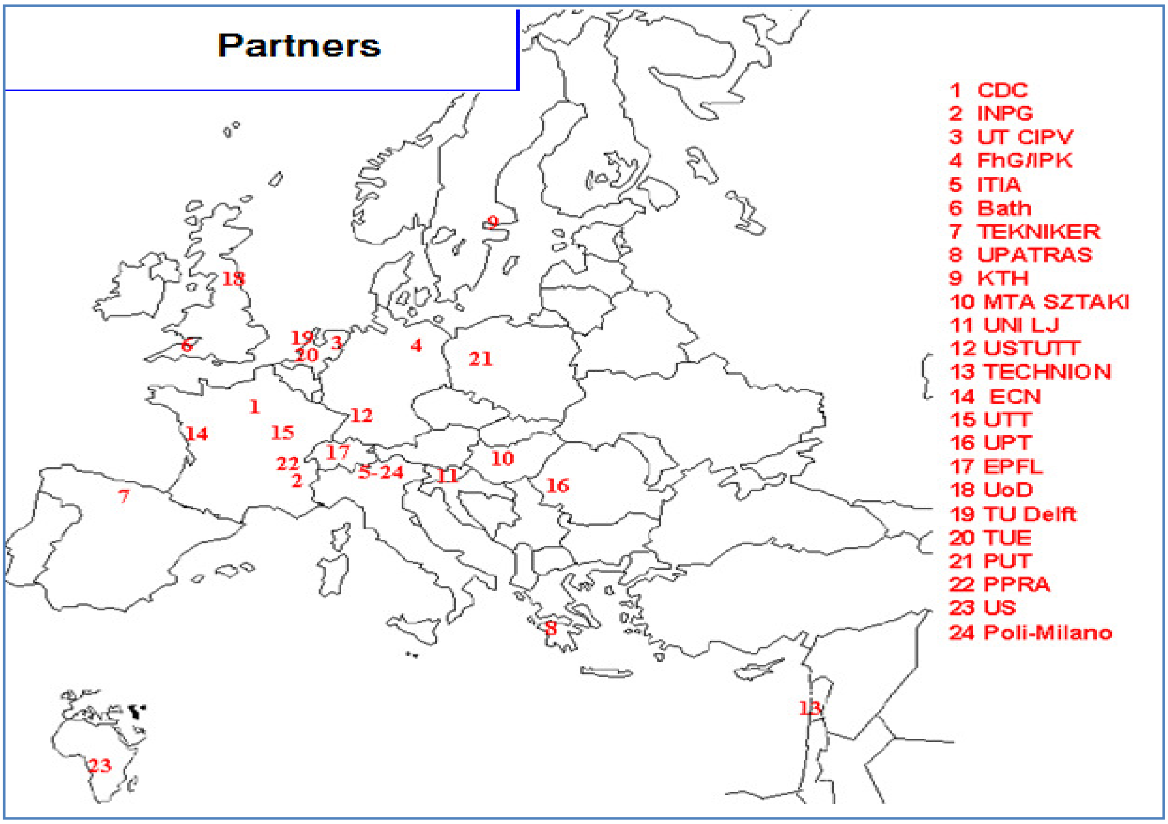


Figure 38: VRL partners

Moreover, The Network has attained the involvement and commitment of industry by involving 30 selected European industries, which play a key role e.g. in:

- Providing industrial viewpoints on relevance and awareness of integration activities and research topics related to production.
- Spreading excellence of joint research outcomes through exploitation of results.

The areas of competences of VRL-KCiP are especially:

- virtual production,
- supply chain,
- life-cycle management,
- interactive decision-aid systems,
- rapid manufacturing

Thus, VRL-KCiP constituted an international platform of excellent research in the various fields of holistic production which provides its know-how in the form of services, tools, and collaborative projects to both industry and academia, supported by IT technologies.

Thus the VRL-KCiP can be considered as a virtual Community of Practice.

Indeed its characteristics are that:

- The network is virtual :
  - it is delocalized,
  - it involves numerous partners from many countries
- It is based on common practices:
  - it deals with the domain of holistic production, and it aims to support the sharing of knowledge about it among researchers
  - the areas of competences are recognized,
  - it clearly emphasizes the necessity for standardisation, i.e. a common language.
- More specifically, VRL-KCiP is a “formal network”, which spans organizations (research laboratories and industrial partners) but is not part of other formal relationships. It has an unlimited number of contributors, but the exchanges are controlled in it by agreement.

XI.2. Theoretical study: impacts of the methodology on VRL’s needs

This study will be based on a few resource documents:

- The VRL Knowledge Axis (Du Preez, 2004).
- The Task105, a document created in the University of KTH for specifying and integrating a KMS so as to support VRL-KCiP. The general needs of the network VRL are particularly emphasized, as well as some use cases to describe the collaborative actions with this KMS (VRL-KCiP, 2006).
- The overview for the Associate members of the VRL – The VRL Joint Programme of Activities, 2nd stage version (Associate members of the VRL, 2004), a document giving an overview on specific planned activities to develop the VRL. This paper is basically divided in three parts:
  - A first part about Knowledge Management, experts identification and constitution of knowledge database, information organization, systems to support them…
  - Another one more on the communication, the dissemination, and the technological transfer among laboratories and industrial partners so as to innovate and discuss.
  - The last one stressing Integration (more focused on the program management, with some milestones and some project tasks)

XI.2.1. The limited scope of the study

The subject of the master thesis was limited to the flow of reification (or information generation). So we dealt with solely with the ways to push information towards users, in organising, easing cognition, enriching and assessing it.

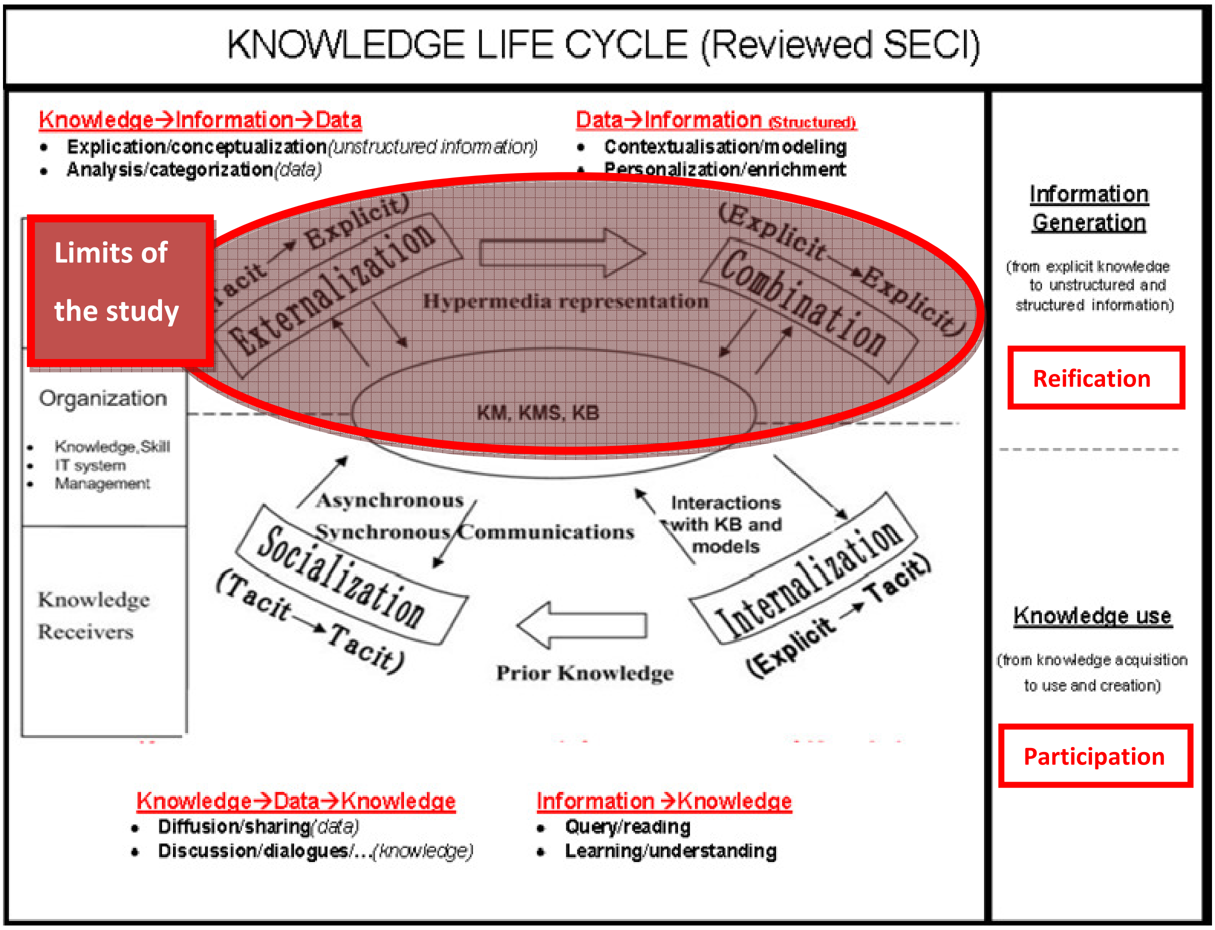


Figure 39: Limits of the study

In the VRL resources, we found a diagram outlining the different steps of Knowledge process (cf. fig.35). After the requirements analysed, the competences of experts and information are identified and gathered. This gathered information becomes resources for managing knowledge and prepare research. The research is then executed, making out new knowledge, which is evaluated and added to the database.

Since we limited the subject to the externalization and combination processes, the flow of “knowledge use” (acquisition, discussion, action, creation) are not in our scope.

So the case study will be limited on:

- ***VRL1.*** the competence analysis and database, when knowledge base and expertise profiles are constituted
- ***VRL2.*** the knowledge management, when knowledge is transformed, modelled, enriched,…
- ***VRL3.*** the research preparation, when the flow of reification push information about people and knowledge towards users, facilitating the future flow of "participation" (or "knowledge uses")

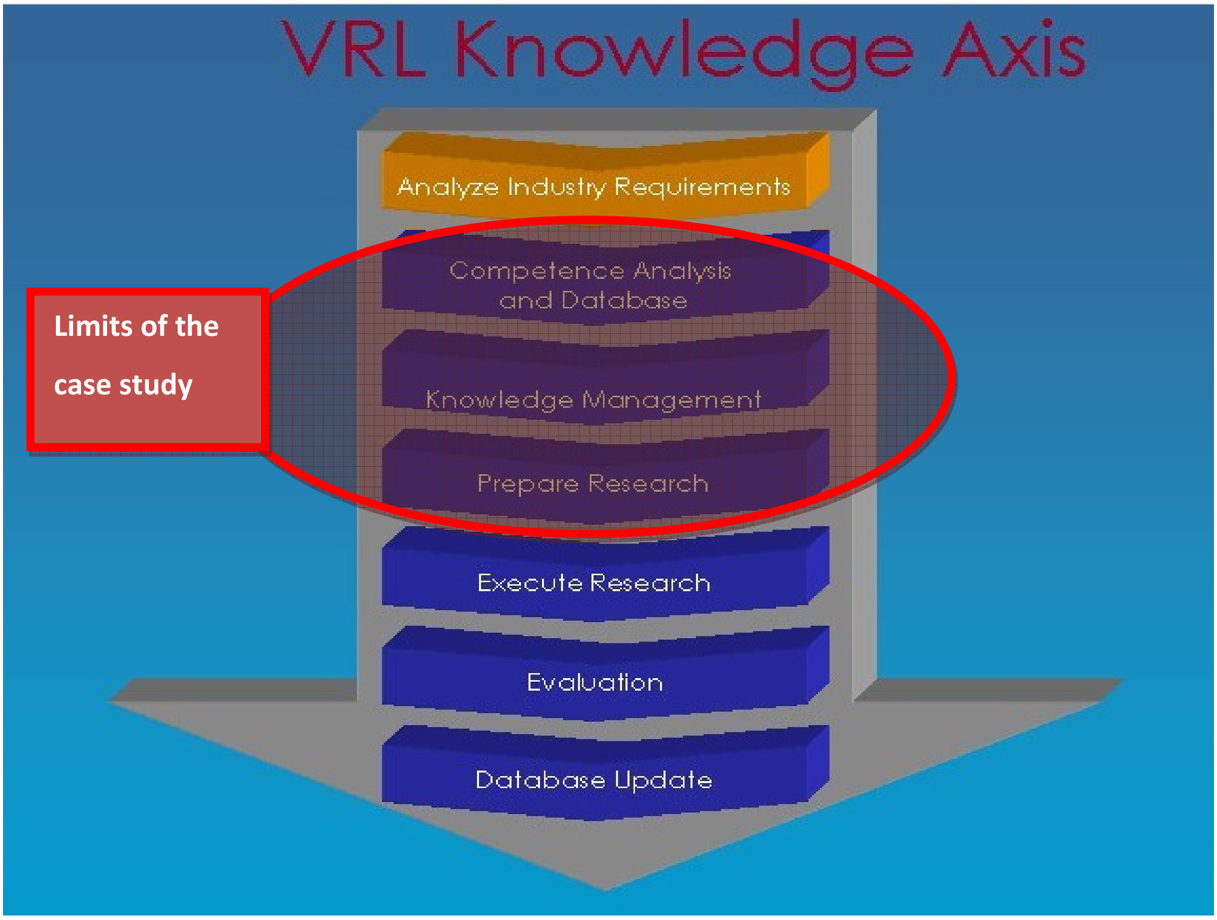


Figure 40: Limits of the case study

XI.2.2. VRL's needs & specifications: Task 105 & Overview for the Associate members of the VRL

Task 105: general needs

According to (VRL-KCiP, 2006), one of the main objectives of the VRL-KCiP is to create a platform enabling the different members of the network to participate in collaborative design projects.

This platform aims at:

N1. ***Facilitating*** VRL members and industry to ***find information and people*** within the VRL network (technology transfer)

N2. ***Facilitating collaborative work*** within VRL-KCiP, and between VRL-KCiP and industrial users

N3. ***Facilitating collaboration and common understanding*** among industrial users

N4. To enable the ***sharing of information in "the right" context*** and ***disseminate the same meaning*** to the different participants.

N5. To ***enable each member to contribute the knowledge related to his own expertise*** as part of a larger whole

N6. To enable each member ***access to and understand in detail*** the part of the content that they need to use

N7. To enable each member ***to understand the scope of the knowledge*** that can be delivered by other partners involved in the network.

Because the subject was limited to information generation in our study, we will be interested solely in the goals related to the flow of reification, i.e. the information push towards users; the possibilities of enriching this information and learning form it, with general and detailed views.

Seemingly, we must thus to evict the goal 2 from this theoretical case study, because it deals more with the collaborative work in the flow of participation (acquisition, use and discussion).

All the other goals (1, 3, 4, 5, 6, and 7) are linked partially to the flow of information generation, so we will consider them and we are going now to see how the proposed methodology fits them.

<u>Overview for the Associate members of the VRL: precise specifications</u>

(Associate members of the VRL, 2004) listed some tasks to implement VRL. We could also use them to show how the proposed methodology can fit the VRL requirements.

In the same way, we have only kept the “jointly executed research activities” part. This par is subdivided in 3 Work Packages: WP1 and WP3 about Knowledge Management and research tools for common use, and WP2 for product models and product development processes.

Within the limits of the case study expressed above, the WP2 is out of it.

Now we are listing below the different points related to our scope:

Work Package 1: Develop new research tools and platforms for common use

- **Task 1.1.** Develop a common KB to support collaborative R&D within the network
  - ***T111.*** Identify design, manufacturing and engineering management lexicon, taxonomy and eventually ontology
- **Task 1.2.** Provide KM and engineering tools
  - ***T121***. Define access and use priorities and constraints
  - ***T122.*** Define specifications on required interfaces to the common toolbox
  - ***T123.*** Agree knowledge capture methodology
  - ***T124.*** Adopt knowledge management toolkit
  - ***T125.*** Adopt data mining toolkit
  - ***T126.*** Specify tools for courses
- **Task 1.3**. Contribute to different content aspect of the common toolbox
  - ***T131.*** Evaluate for each partner key competencies, existing expertises and areas of interest
  - ***T132.*** Define requirements for the tools
  - ***T133.*** Analyse and evaluation of the toolboxes used within the network
  - ***T134.*** Definition of uses scenario and typical workflows

Work Package 3: Knowledge Management

- ***KM1.*** Collect and formalize knowledge about production processes, design and virtual prototyping…
- ***KM2.*** Define knowledge management methodologies and tools for sharing knowledge and applications for demonstration inside the network

XI.2.3. Analysis of VRL's needs and specifications

In order to clarify all these needs and specifications and before studying how they can be solved partially by our propositions, we tried to sort them in the following scheme.

The horizontal axis represents the level of specifications (from knowledge framework to high level specifications), whereas the vertical axis focuses more on a process's point of view, where Competences analysis and database are the inputs, KM is the process which adds value to these inputs, and research preparation are the outputs (when people use KM to find information and people in order to collaborate and innovate).

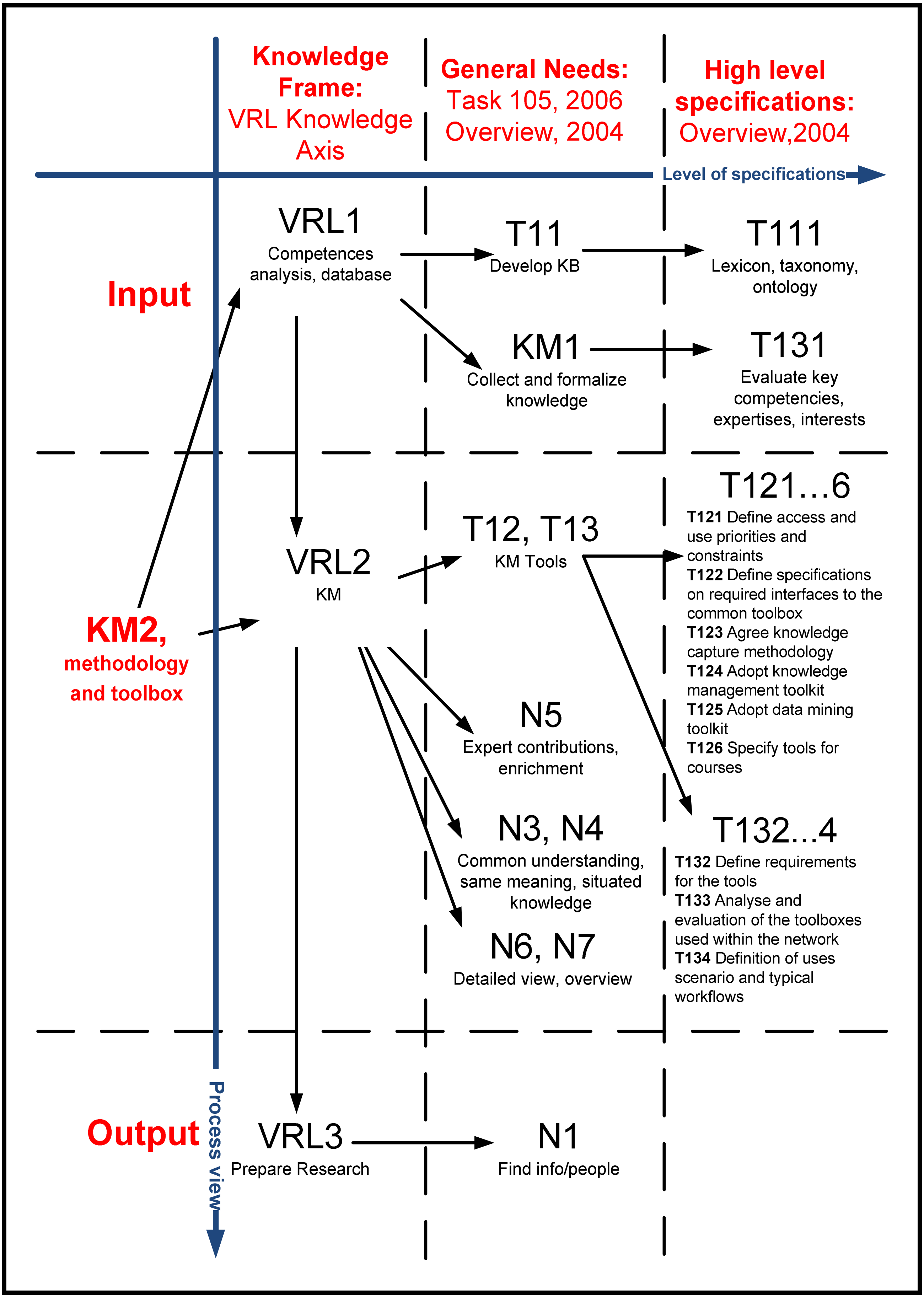


Figure 41: Analysis of VRL's needs and specifications

XI.2.4. Responses of proposed methodology and tools for VRL

Methodology and Toolbox: Response to KM2 specification

We can observe that the development of this master thesis, i.e. the proposed methodology and the choice for tools, seems to respond obviously to the specification KM2, in bringing out some suggestions for sharing knowledge inside the network VRL.

The different steps and components of the Chapter IX of our thesis are then spread into the granular levels of VRL's needs, in the part of "Competence analysis and database" (VRL1) and "KM" (VRL2).

Observation about the specific members of VRL-KCiP

The VRL network is a specific Community of Practice. Indeed, the VRL's members can be either researchers or industrial partners.

That triggers off different profiles, and the management of the user's interactions and authorizations will be affected by this fact.

Indeed, we can assume that the industrial people are certainly more interested in results of research, and so they are more consumers of knowledge, and more passive members.

By contrast, the researchers from VRL are more implied in the generation of the knowledge base, writing and modifying some documents, updating the information in the system. So they are more active members in the reification process.

How the methodology and the toolbox fit VRL's needs and specifications?

We used again the previous scheme about the analysis of VRL's needs and specifications, and we added the different points of the methodology and the toolbox which can bring a solution to VRL (cf. Fig.37).

To do this, we compared the Fig.36 and the summary of Table 17.

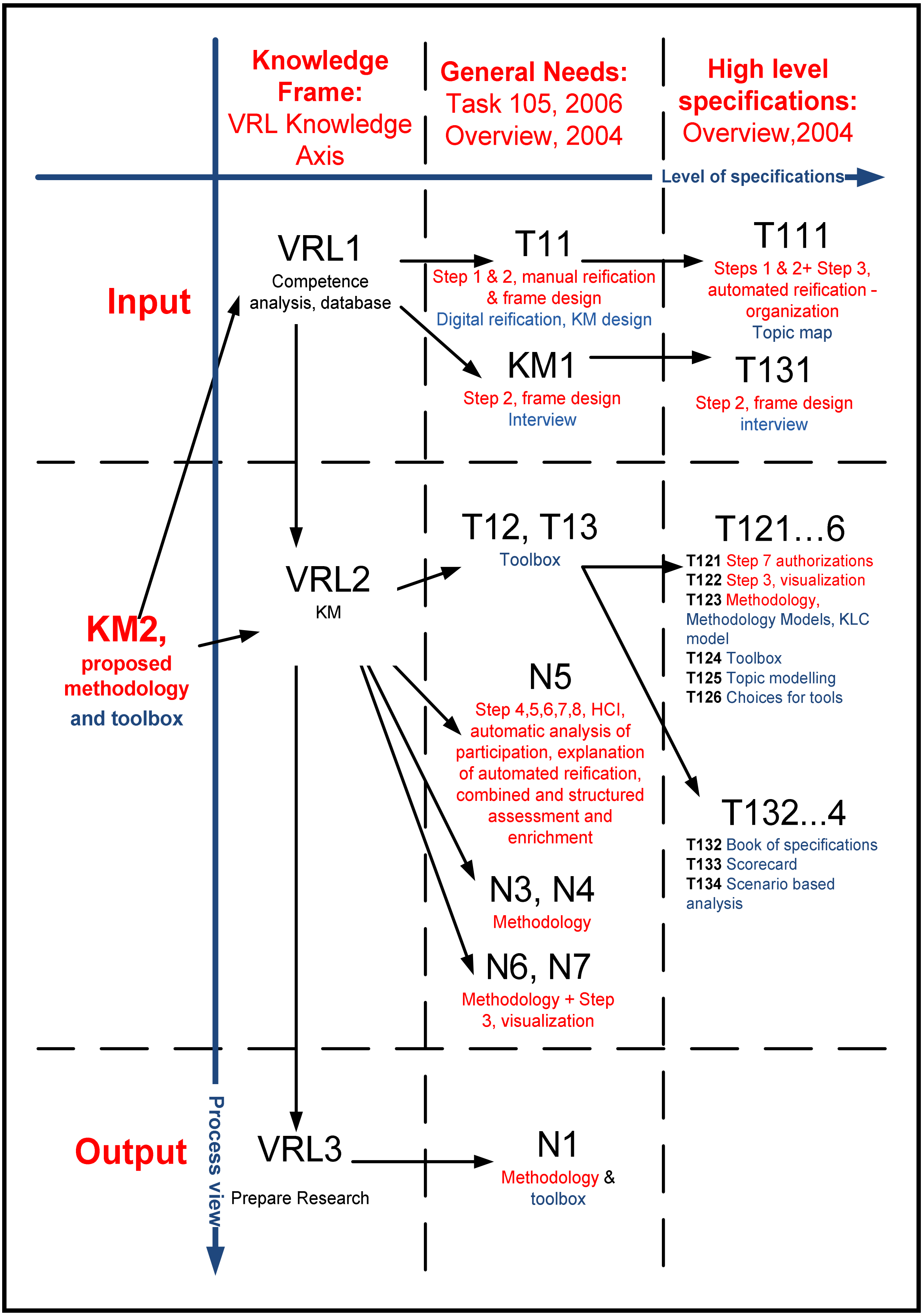


Figure 42: Methodology and toolbox versus VRL's needs and specifications

XI.3. Conclusion

Seemingly, the proposed methodology and the chosen toolbox can respond to some general needs and some specifications of the Network of Excellence VRL-KCiP, a virtual Community of Practice, within the given limits of our subject.

It is fundamental to understand the methodology in its context: because we limited our study to the flow of reification, we did not focus on the participation and the use of information to collaborate and create new knowledge. It is why we did not study the documents and the specifications about collaborative work.

The methodology and the toolbox have especially an impact at the level of Knowledge Management in VRL and its inputs, i.e. the gathering of information about people (competences analysis), the constitution of a Knowledge Base, then the organization, the modelling and the collaborative enrichment of information.

Indeed, they allow to push information towards the different users, in diffusing a situated knowledge and common practice, in order to prepare the research work.

This case study is only theoretical, because of a lack of time and means. A further work would to measure practically the efficiency and the effectiveness of the methodology and the toolbox in VRL or in a similar network.

Conclusions and trends

CONCLUSIONS

Summary of the work done

We summed up the different steps of our work in the following scheme.

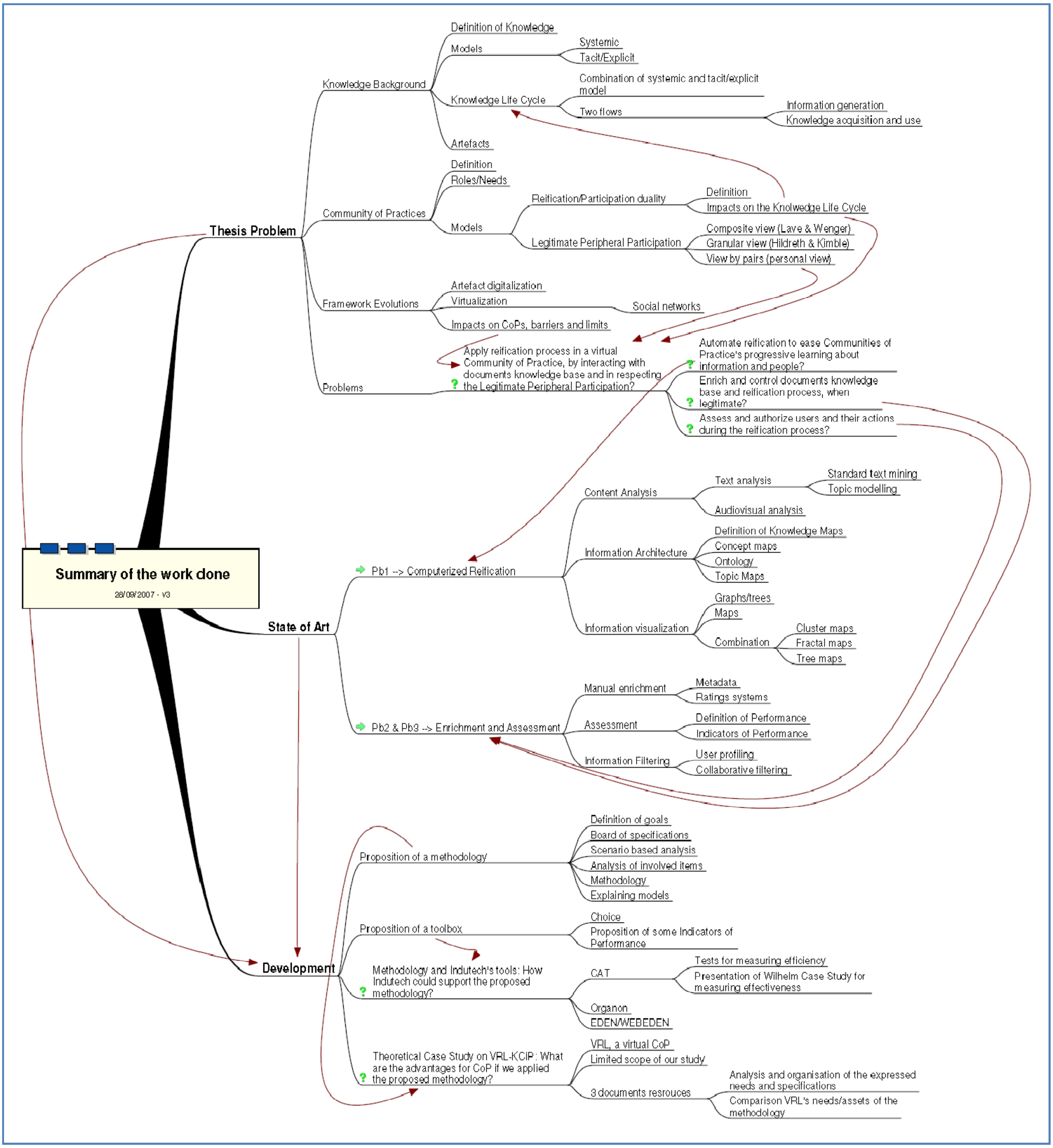


Figure 43: overview on the work done

Discussion about the work done

A current, theoretical and limited Thesis Problem

In this study, we have attempted to solve the application of information generation in particular knowledge networks, called communities of Practice, in following the essential principle of Legitimate Peripheral Participation in an evolving context.

This problem comes from the observations of recently recognized knowledge structures, the CoPs, and their formalized models of functioning, as well as the social networks and their limits in a virtual framework.

Moreover, the raised problem is a theoretical problem, limited to the CoP's structure and limited by the angle of view. Indeed, we decided to limit the field of research to the flow of reification in the knowledge background, and to exclude the flow of participation, where knowledge is internalized and socialized.

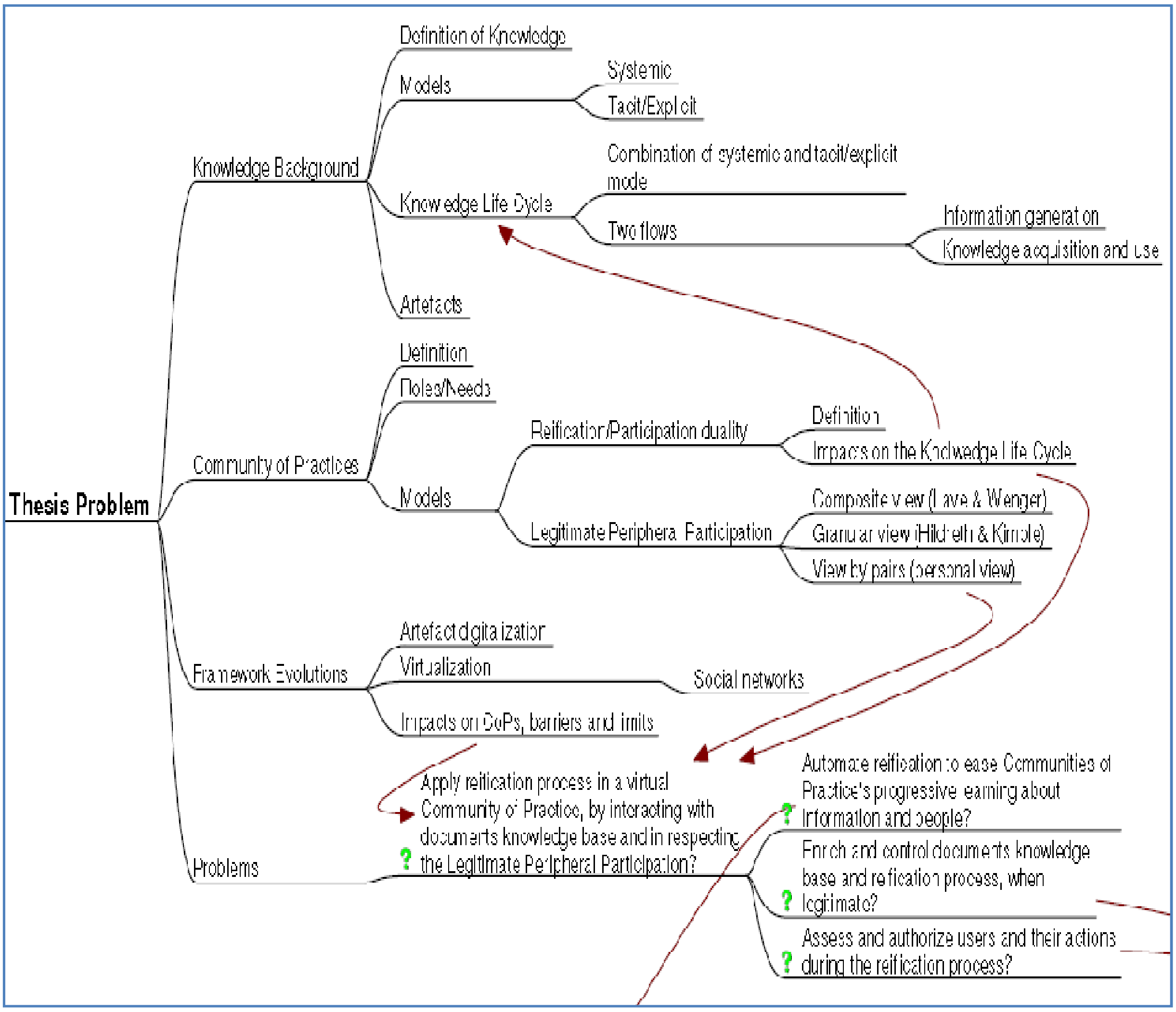


Figure 44: Summary of the thesis problem

A practical and oriented State of art

Given the divided aspect of the thesis problem, we have led the state of art in focusing on the techniques and methods which could potentially respond to them.

The state of explore a broad field, from information extraction to visualization, through the way to enrich and assess manually or automatically extracted contents. That provides also a better understanding of the different investigated domains, so as to prepare the development and propose a relevant methodology.

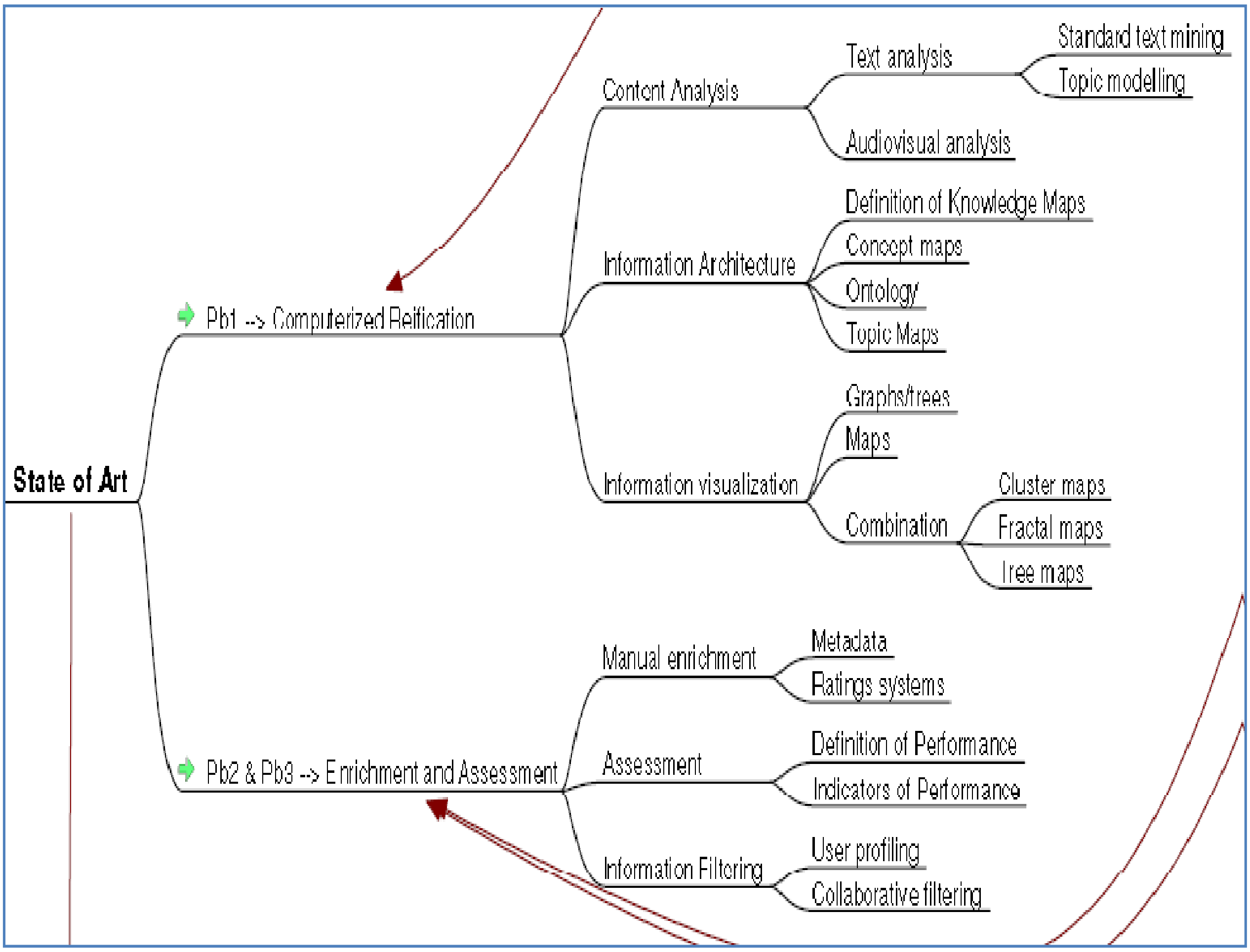


Figure 45: Summary of the State of Art

Development, between propositions and verifications

Finally, we tried in the development to solve an organizational and functional problem with a technical state of art. Thus, the study was a real problem of KM, mixing organizational management and Information Systems.

We have built and proposed a solution, composed by a methodology, some explaining models and a toolbox. To do that, we adopted an "engineering" approach, in defining the goals of the study, in making a board of specifications, then in analysing some other factors, with the aid of a scenario-based analysis and the analysis of the different involved items.

So as to put our propositions in a concrete framework and open more the cogitation about the range of the methodology, we studied:

- How Indutech and its tools could support the proposed methodology. We have especially tested the possibilities of a particular topic modelling tool, CAT.
- How the methodology would fit the needs and the specifications of a robust virtual CoP, VRL-KCiP, if it was applied.

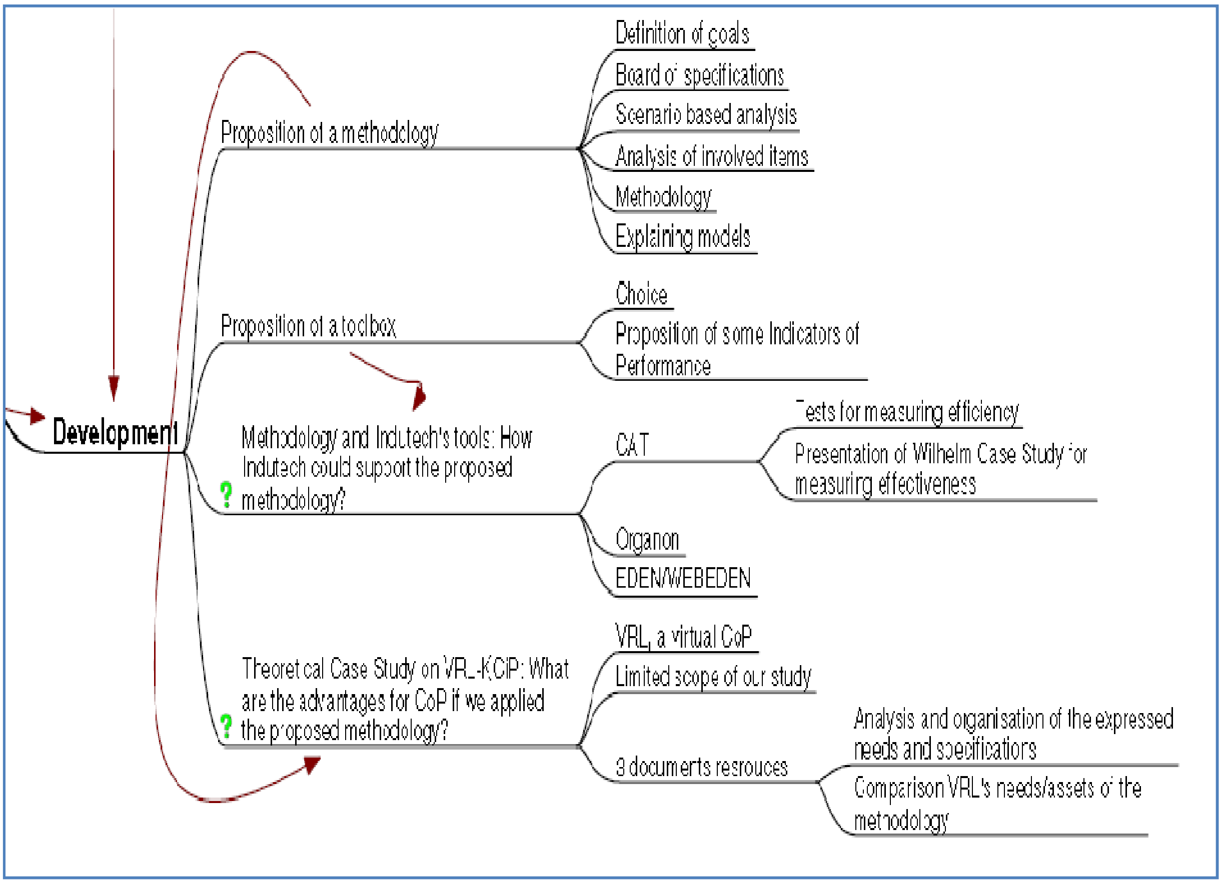


Figure 46: Summary of the development

PERSPECTIVES

Possible continuation for the project

Methodology and Case Study

The" theoretical Case Study" allows only to verify the adequacy between the needs of a real virtual CoP and our assumptions.

Unfortunately, we did not have enough time to verify the proposed methodology with a practical Case Study.

So it would be interesting and necessary to make such a practical study, in order to measure the robustness of the suggested models and complete them.

Indutech's tools

The short study about the software products of Indutech could be continued. The practical case study could be for instance tested in using CAT, Organon and EDEN as the base for supporting the methodology.

Further researches

Trends towards the flow of participation

We have limited our subject to the flow of reification, i.e. a part of the knowledge life cycle which rules the functioning of CoPs. In further works, it would be necessary to make the study of the flow of participation, i.e. how people retrieve information, use it and interact in order to innovate.

The field of information retrieval, search engine ergonomics, and collaborative tools and platforms should be investigated.

Furthermore, this study of the "other side" of the Knowledge Life Cycle would provide a means to measure the interest of the propositions of this thesis, and emphasized the impacts of an automated and interacted reification for helping CoP's members to acquire, exchange and create knowledge.

Finally, because reification and participation are not really undividable, a model needs further to be proposed to link the automated and enriched reification with its use, its exchanges and the creation of new knowledge by people.

FINAL REMARKS

This research work in Indutech provided me many and many benefits, both at an academic level and at a personal level.

On the one hand, I had occasion to investigate the domain of the research and its functioning. I needed to follow a new way of thinking and arguing, where one build oneself his problems before contributing and suggesting a solution. That was a little bit different from that one we learned in engineering school, where the problems are generally already formulated.

To fulfil this master degree, I learned and acquired some knowledge in the fields of Knowledge Management and Information Systems. If the subject was limited to the Communities of Practice, these researches have me consider Internet and my use differently, especially for the domain of information organization and retrieval as well as social networks.

On the other hand, I discovered a fabulous country and many very kind persons. I worked and took fun with people who have other cultures (a mix between English, Afrikaans, Xhosa, Zulu,…), other languages (eleven official languages). This environment was very exciting and enriching.

I have also visited many different places in South Africa, from the Western Cape to the Eastern Cape, through the little Karoo and the Drakensberg, which provided me a lot of fantastic sceneries and satisfy my traveller's spirit.

To conclude, this experience demonstrated me that I enjoyed this research work and I would like continue in the domain of research, hopefully in making a PhD, in learning more about Knowledge Management and exploring some other fields. I am also sure to come back in South Africa, because I fell in love with this country!

Appendices

TABLE OF APPENDICES

Appendix 1. Master Scope Definition & Initial Planning 130

Appendix 2. Classification of Communities of Practice 132

Appendix 3: Typologies of knowledge maps 133

Appendix 4: Visualization methods 136

Appendix 5. Survey on visualization tools................................ 137

Appendix 6. Survey on collaborative filtering tools.................. 138

Appendix 7. Flow based-model step by step............................. 139

Appendix 8. Results of tests on CAT 143

Appendix 9. CIRP Case Study: Overview of Results 144

APPENDIX 1. MASTER SCOPE DEFINITION & INITIAL PLANNING

Philippe's Project

- Philippe's Project
  - 1. Research Objectives & Planning
  - 2. Literature Study
  - 3. Development
  - 4. Research Report/Thesis
  - 5. Research Closure
  - 6. External Inputs

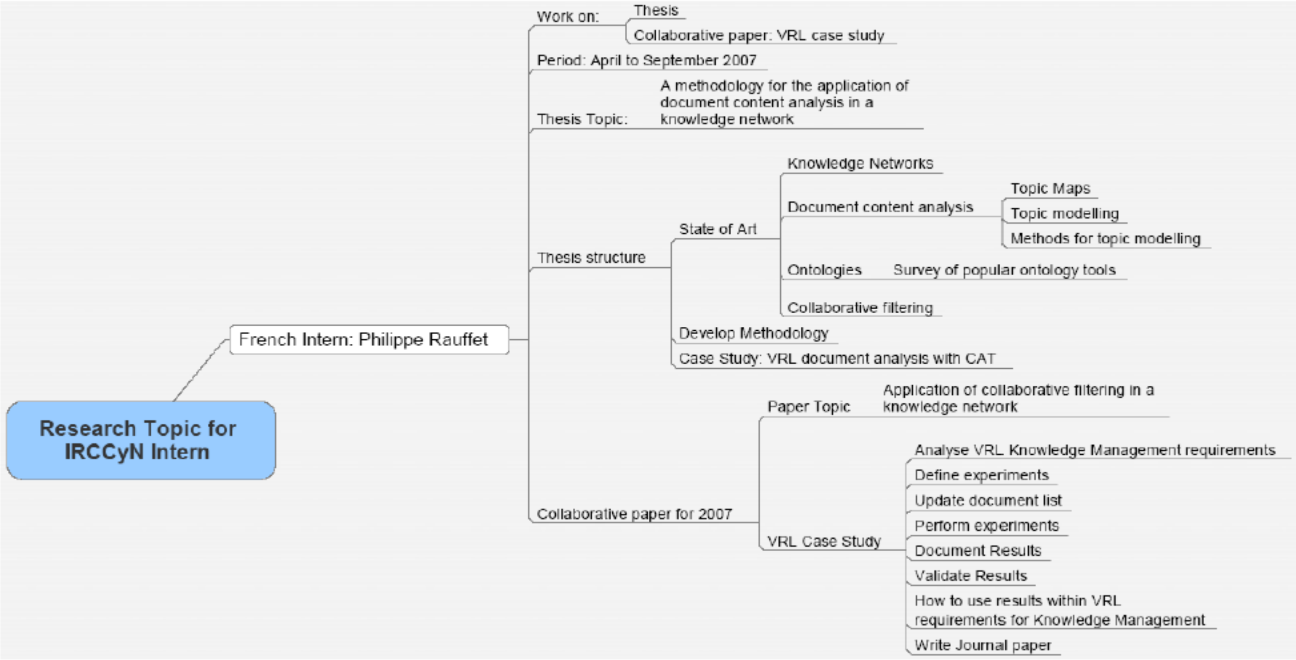
"a methodology for the application of document content analysis in a knowledge network"

"Application of collaborative filtering in knowledge network"

Research Objectives & Planning

We defined on Friday, April 13th the main orientations of my research work, in relation with Indutech business and VRL project.

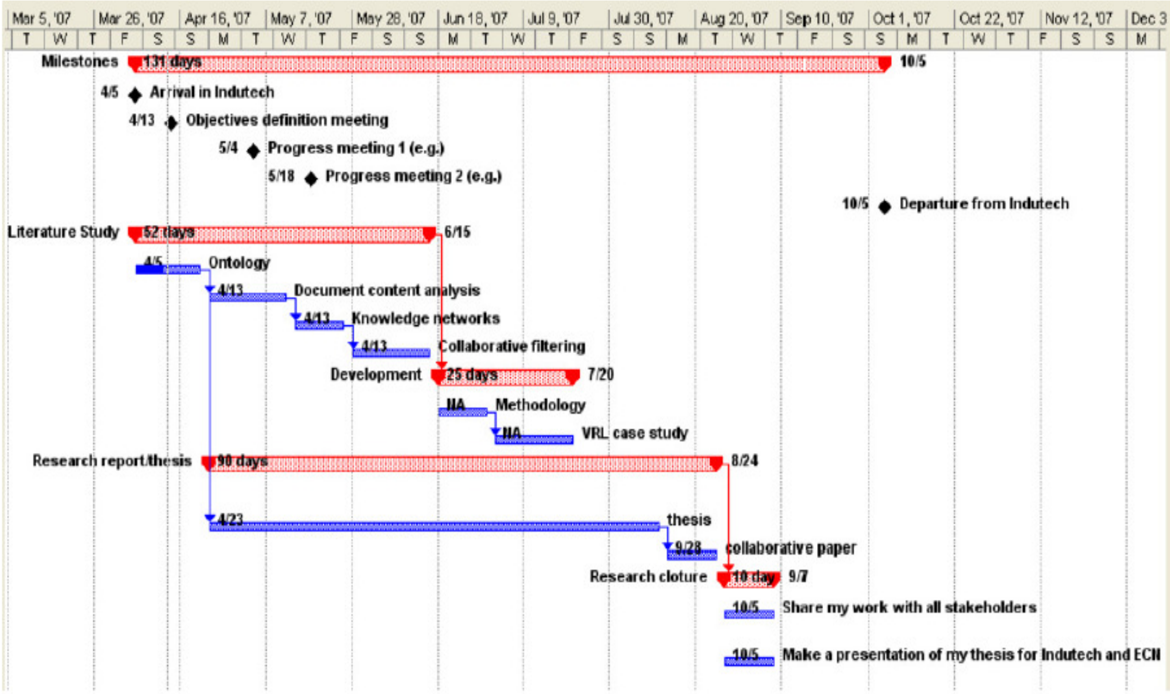
Scope definition



Screen clipping taken: 13/04/2007; 12:08

Planning

		Task	Period	Start	End	Predecessor
1		**Milestones**	**131 days**	**Thu 4/5/07**	**Fri 10/5/07**	
2	✓	Arrival in Indutech	0 days	Thu 4/5/07	Thu 4/5/07	
3	✓	Objectives definition meeting	0 days	Fri 4/13/07	Fri 4/13/07	
4		Progress meeting 1 (e.g.)	0 days	Fri 5/4/07	Fri 5/4/07	
5		Progress meeting 2 (e.g.)	0 days	Fri 5/18/07	Fri 5/18/07	
6		Departure from Indutech	0 days	Fri 10/5/07	Fri 10/5/07	
7		**Literature Study**	**52 days**	**Thu 4/5/07**	**Fri 6/15/07**	
8		Ontology	2.4 wks	Thu 4/5/07	Fri 4/20/07	
9		Document content analysis	3 wks	Mon 4/23/07	Fri 5/11/07	8
10		Knowledge networks	2 wks	Mon 5/14/07	Fri 5/25/07	9
11		Collaborative filtering	3 wks	Mon 5/28/07	Fri 6/15/07	10
12		**Development**	**25 days**	**Mon 6/18/07**	**Fri 7/20/07**	**7**
13		Methodology	2 wks	Mon 6/18/07	Fri 6/29/07	
14		VRL case study	3 wks	Mon 7/2/07	Fri 7/20/07	13
15		**Research report/thesis**	**90 days**	**Mon 4/23/07**	**Fri 8/24/07**	
16		thesis	16 wks	Mon 4/23/07	Fri 8/10/07	8
17		collaborative paper	2 wks	Mon 8/13/07	Fri 8/24/07	16
18		**Research cloture**	**10 days**	**Mon 8/27/07**	**Fri 9/7/07**	**15**
19		Share my work with all stakeholders	2 wks	Mon 8/27/07	Fri 9/7/07	
20		Make a presentation of my thesis for Indutech and ECN	2 wks	Mon 8/27/07	Fri 9/7/07	



Screen clipping taken: 13/04/2007; 16:31

Description of tasks

Literature study

First I should study literature about :

- **Knowledge Networks** : I need to understand these network, their needs, their different sharing exchanges and communication, and how they work. I would be able to use it to characterize VRL network, where every people is linked to another, but with unequal links. I need also find the requirements of VRL Knowledge Management System. Maybe I should have to look for Ecosystem network, to define more precisely VRL and its functioning.

- **Document content analysis** : I will have to do a state of art on tools of concepts extraction and representation. So I will look for tools and methods for mapping and modelling conceptual framework. For example I could base on visual search engine tools for representation and CAT for extraction.

- **Ontology** : Although ontology is different from topic maps (because it add a formal machine-understandable constraint), it should be interesting to explore this domain to find other tools or methods to model topics. Maybe see in details Protégé and GraphViz….

- **Collaborative filtering** : It is means to suggest to linked concepts to user who are looking for a specific concept. For example, tags, deli.ci.ous, youtube,…

Development

- **Methodology** : After this bibliographical work, I should develop a methodology, from the extraction of "concepts" in documents to the topic modelling, by using tools and methods I would have seen previously.

- **VRL case study** : I will apply my previous work on the VRL network, for example in testing a corpus into CAT software and attempting to modelling output conceptual framework (maybe also test if CAT works with a big corpus and with a language-mixed corpus...).

Research report/thesis

I will have to write two documents, a report and a paper.

- **Thesis Topic** : "a methodology for the application of document content analysis in a knowledge network"
(First Literature Study about KM, State of art about document content analysis and collaborative filtering, Developped methodology for modelling topics, test on VRL, Conclusion and Trends,...)

- **Paper Topic** : "Application of collaborative filtering in knowledge network"
(more focused on the application of methodology on VRL)

Research closure

- **Academic** : I will present my work in Nantes in October 2007.

- **Stakes** : Create a collaboration with VRL KCiP, present during general VRL conference, in one year, Indutech's tools to attempt to implement them in this research network?

APPENDIX 2. CLASSIFICATION OF COMMUNITIES OF PRACTICE

(Archer, 2006)develops and characterizes his classification in the following table:

	Community of Practice			
Characteristic	*Internal*	*Network Organization*	*Formal Network*	*Self-Organizing Network*
Type of Knowledge	Product, service (technical), management skills, processes	Product, service (technical), management skills, processes	Management skills, processes; operational, product knowledge	Management skills, processes; operational, product knowledge
Desired Objective or Outcome	Innovations in products, services, improved management practices	Innovations in products, services, improved management practices	Improved management practices, products, services	Improved products, services, management practices
Funding	Internal	Shared	Shared	Voluntary
Intellectual Property	Internal	Shared by formal agreement	Controlled by the network	Shared by agreement
Management	Internal	Managed jointly as component of organizational agreement	Externally managed	Externally managed
Professional Expertise	Internal	Shared by formal agreement	Shared by agreement	No agreement
Dispute Resolution	Internal management	Legally resolved	Withdrawal	Withdrawal
Potential Knowledge Contribution	Unlimited from internal sources with need to know	Limited by formal agreement	Determined by members; No min. or max. limit	Determined by members; No min. or max. limit
Common Benefits	Developing and sharing formal best practices, learning and sharing tacit and explicit knowledge, benchmarking, innovations in management, operations, and processes			
Potential Gain – Shared Knowledge of:	Innovations in products, services	Innovations in products, services, shared access to IP	Innovations in management practices, innovations in products, services	Innovations in practices, understanding and innovating products, services
Common Problems	Unpredictable payback, initiating and maintaining interest, building and maintaining trust, encouraging steady flow of information and knowledge among participants, divergence of objectives, lack of common participant language (natural and/or professional), ensuring payback to all participants			
Potential Problems	Reorganization may be required to improve knowledge sharing and learning	Limitations of formal agreement	Ensuring knowledge contributions from all members	Unknown value of knowledge communicated; Hard to reach contributors
Remediation of Operational Problems	Attention from moderator or manager(s)	Attention from moderator, manager(s), or legal resort	Attention from moderator	Targeted attention from membership
Some Published Examples	IBM (Gongla & Rizzuto, 2001), AMS (Wenger & Snyder, 2000), P&G (Sakkab, 2002)	Toyota (Dyer & Nobeoka, 2000), Biotech firms (Oliver, 2001) Sematech (Davenport, 1997)	Critical Emergency Ops Link (D'Amico, 2002), ASAE Futures Scan (Mason, 2001)	Usenet groups (Faraj & Wasko, 2001), Democracy Online (Cashel, 2002)

APPENDIX 3: TYPOLOGIES OF KNOWLEDGE MAPS

This typology is given by (Greenwood, et al., 2006):

Concept map: Node-link structure in which nodes denote concepts and links show the relationship between these concepts

- Express a particular person's knowledge about a given topic in a specific context; explore prior knowledge and misconceptions; problem-solving tool, shorthand form for organising and sequencing ideas

Mind Map /Idea map: Begin with one central theme or concept and radiate hierarchically outwards as sub-concepts/additional ideas emerge

- Note-taking technique; a way to capture and reflect the processes in the brain; used for training scheme manual

Concept circle diagram: Labelled circles which may be inclusive, exclusive, and/or overlapping to show the relationship between concepts

- Show the existing and desired relationship between concepts, organisations, departments etc.; organise ideas into categories

Semantic map: Similar to mind, idea and concept maps; 6 different types: definitional, assertional, implicational, executable, learning and hybrid networks

- Used for artificial intelligence and machine translations but also previously used in philosophy and languages (see concept map)

Cognitive map: Mapping of thoughts a person has about a particular situation or problem of interest; from concept mapping to word webbing

- Refer to the mental models that people uses to perceive, contextualise, simplify, and make sense of otherwise complex systems

Process map: Blocks of activities or tasks sequenced in a logical way to achieve a specific goal/ objectives include factors like timescale, resources etc.

- Define the sequence of tasks which link the actions of people within & across functions in order to achieve a specific goal

Social mess map / Cross boundary causality map: Problems are situated in a tangled mess of causal factors that cross the boundaries of stakeholders, processes, industries and social arenas

- Summarise a particular group's understanding of the problems, causes, influences and relevant data about the mess

Conceptual map: Graphical means to compose concepts and conceptual relations

- Used as a communication language between individual and computer systems

Knowledge flow map: High-level knowledge models in a transparent graphical form

- Used to map and measure relationships and flows between people, groups, organisations and information/knowledge processing entities

Causal map: Structure of people's causal assertions and acquisition of consequence following those assertions

- To explore beliefs of individual or groups in order to establish cause & effect relationships; 'what-if scenarios'

Ontology: Working model of entities and interactions either generically, or in some particular domain of knowledge or practice; 3 types: domain-oriented, task-oriented and generic

- Tool in searching all of the available information in a given field

Petri net: Abstract, formal model of information flow; consists of places, transitions and directed arcs; 2 types: stochastic and generalised stochastic

- In search for natural, simple, powerful methods for describing and analysing the flow of information and control in systems

Cluster Vee diagram: Road map showing a route from prior knowledge to new and future knowledge

- It has been developed to help students studying science make explicit essential elements to constructing scientific knowledge

Thesauri: Set of concepts in which concept is characterised by hierarchical, synonymous, horizontal, and other relevant relations

- Used in retrieval system and modern information (e.g. Web, bibliographic records etc.)

Visual thinking network: Meta-cognitive and knowledge representation strategy that encourages the user to integrate multiple ways of thinking that inform concept formation

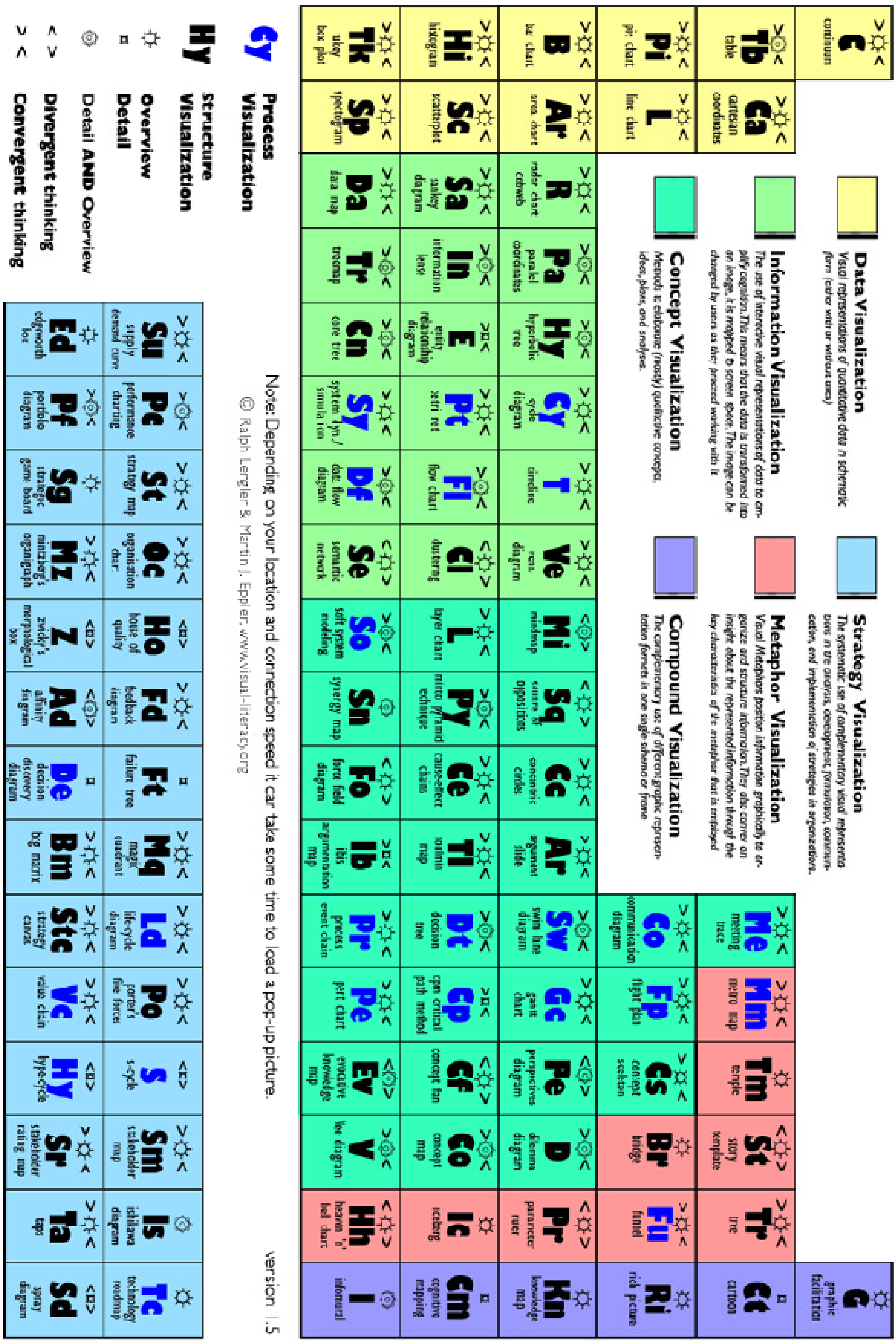
- A technique by which the user can represent, organise and revise their meaning-making of knowledge by grouping and linking symbolic and pictorial visualisations into a coherent whole

Topic map: Electronic versions of back-of-book indices

Perceptual map: Simply way of taking complicated results from research surveys and presenting them on a clear and informative map

- Standard way to visually summarise the dimensions that customers use to perceive and judge products and identify how competitive products are placed on those dimensions

APPENDIX 4: VISUALIZATION METHODS



APPENDIX 5. SURVEY ON VISUALIZATION TOOLS

Sources are from:

http://www.cs.umd.edu/class/spring2005/cmsc838s/viz4all/viz4all_a.html

	Tree/Graphs	Maps	Observations	Neighbourhood	Overview	Filtered view	Navigation
NetVis	X (static)		Graph	nodes	limited	No	No
Glass Engine	X (static)		Tree	Nodes	limited	No	No
InXight - 2001 Santa Clara Housing Market	X (static)		Tree	Nodes	limited	No	No
InXight - SuperBowl predictor	X (static)		Tree	Nodes	limited	No	No
InXight - top 100 movies	X (static)		Tree	Nodes	limited	No	No
State Cancer Profiles	X (static)		Tree	Nodes	limited	No	No
PieSoy: Shakespeare Plays	X (dynamic)		Tree	nodes	limited	Browsing	Yes
Visual Thesaurus	X (dynamic)		Tree	nodes	Limited	Browsing	Yes
Lexis/Nexis map of available sources	X (dynamic)		Tree	Nodes	Limited	browsing	Yes
Dynamic Choropleth Maps		X	Map	Geographical	Yes	No	Yes
Fannie Mae: DataPlace		X	Map	Geographical	Yes	No	Yes
Fundrace: Money Map		X	Map	Geographical	Yes	No	Yes
Microsoft Netscan - Usenet crosspost	X (dynamic)		Cluster	nodes	Yes	expanding/	Yes
They Rule	X (dynamic)		Cluster	nodes	Yes	expanding/	Yes
CiteSpace	X (dynamic)	Overview,	Cluster	node, colour	Yes	expanding/	Yes
Live Plasma	X (dynamic)	Overview,	Cluster	node, colour	Yes	expanding/	Yes
Touch Graph: Google results	X (dynamic)	Overview,	Cluster	nodes, colour	Yes	expanding/	Yes
Hive Group - Amazon product search	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
Hive Group - Google results	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
Hive Group - iTunes top 100 songs	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
Hive Group - World Data	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
News Is Free: Top News	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
NewsMap	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
Peet's Coffee: Coffee Selector	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
SmartMoney: Map of the Market	Hierarchy	X	Treemap	colour	Yes	by hierarchy	No
Grokker	X	X	Fractal view	colour	Yes	by "zooming"	Yes
Fractal Edge	X	X	Fractal view	colour	Yes	by "zooming"	Yes

APPENDIX 6. SURVEY ON COLLABORATIVE FILTERING TOOLS

This short survey can be found on http://en.wikipedia.org/wiki/Collaborative_filtering

In commercial systems

Commercial sites that implement collaborative filtering systems include:

Amazon Barnes and Noble half.ebay.com Hollywood Video	iLike - music Last.fm - music LibraryThing - books Musicmatch Netflix	TiVo Thoof Minekey - Online recommendations service. StumbleUpon - websites

In non-commercial systems

Non-commercial sites that implement collaborative filtering systems include:

AmphetaRate - RSS articles Everyone's a Critic - movies GiveALink.org - websites Gnomoradio - free music	iRATE radio - free music Moonranker - music, movies, and books MovieCritic - movies, MovieLens - movies	Music Recommendation System for iTunes - music Musicmobs - music MyStrands - music Rate Your Music - music

Software libraries

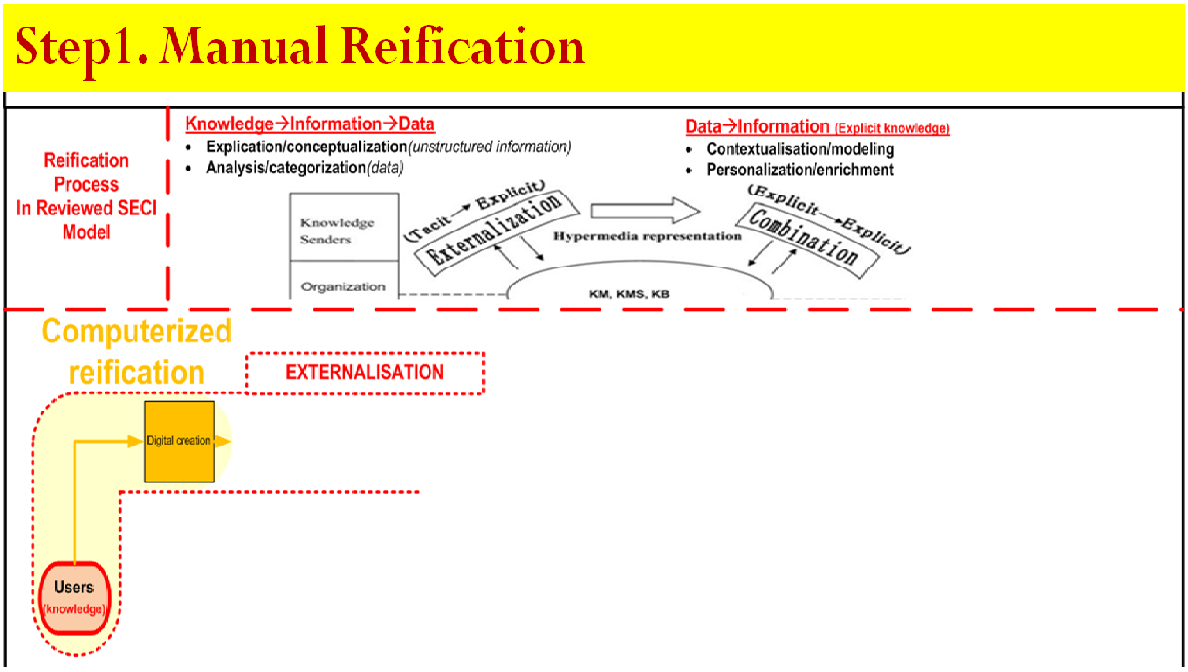
There are also software libraries which enable a developer to add collaborative filtering to an application or web site:

Taste - open-source, Java Cofi - open-source, Java CoFE - open-source, Java ColFi - open-source, Java	RACOFI - open-source, Java SUGGEST - Free, written in C. (A library, not open source.) Rating-Based Item-to-Item - public domain, PHP Vogoo PHP Lib - open-source, PHP	consensus - open-source, Python C/Matlab Toolkit for Collaborative Filtering - open-source, Matlab, C Fast Maximum Margin Matrix Factorization - Matlab/Octave

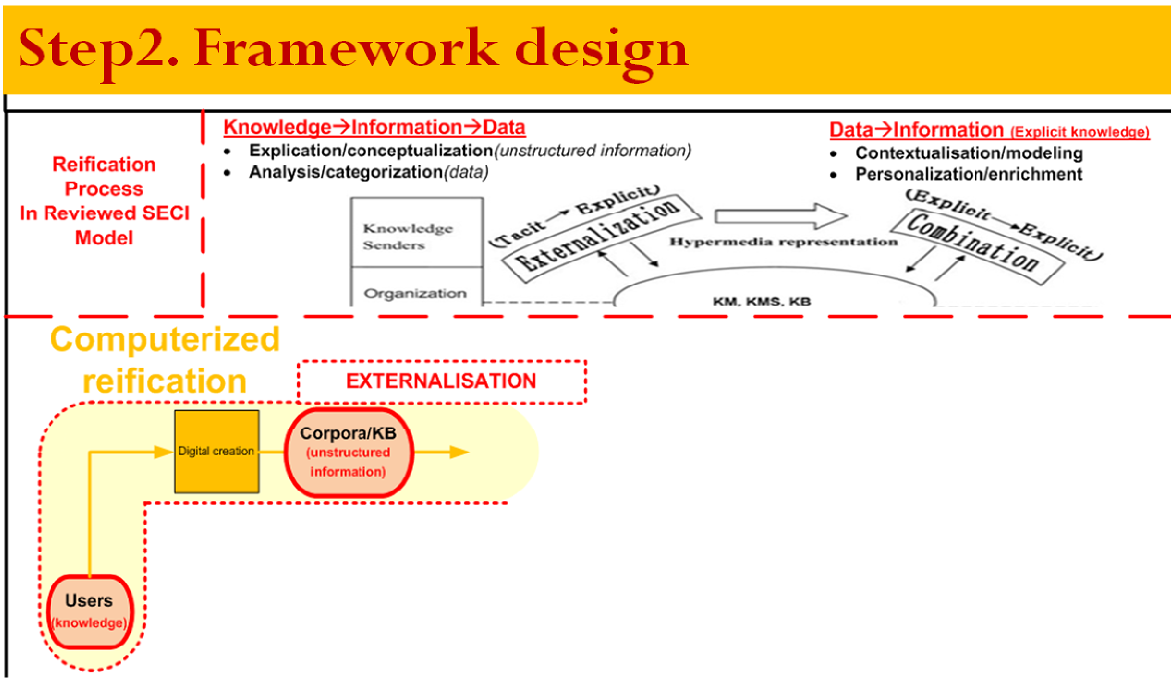
APPENDIX 7. FLOW BASED-MODEL STEP BY STEP

We based our model on the reification flow of the knowledge life cycle presented on fig.6, showing the externalization and the combination of knowledge expressed by the Nonaka's view.

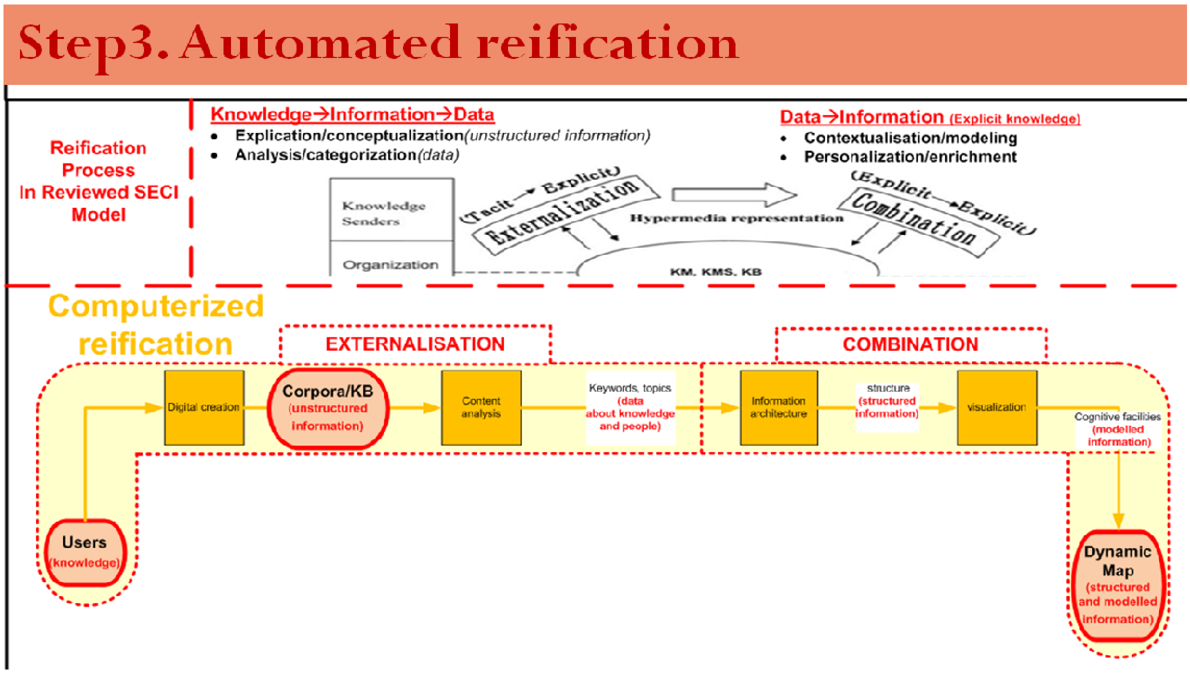
The first step is the manual reification, where users create digital artefacts.



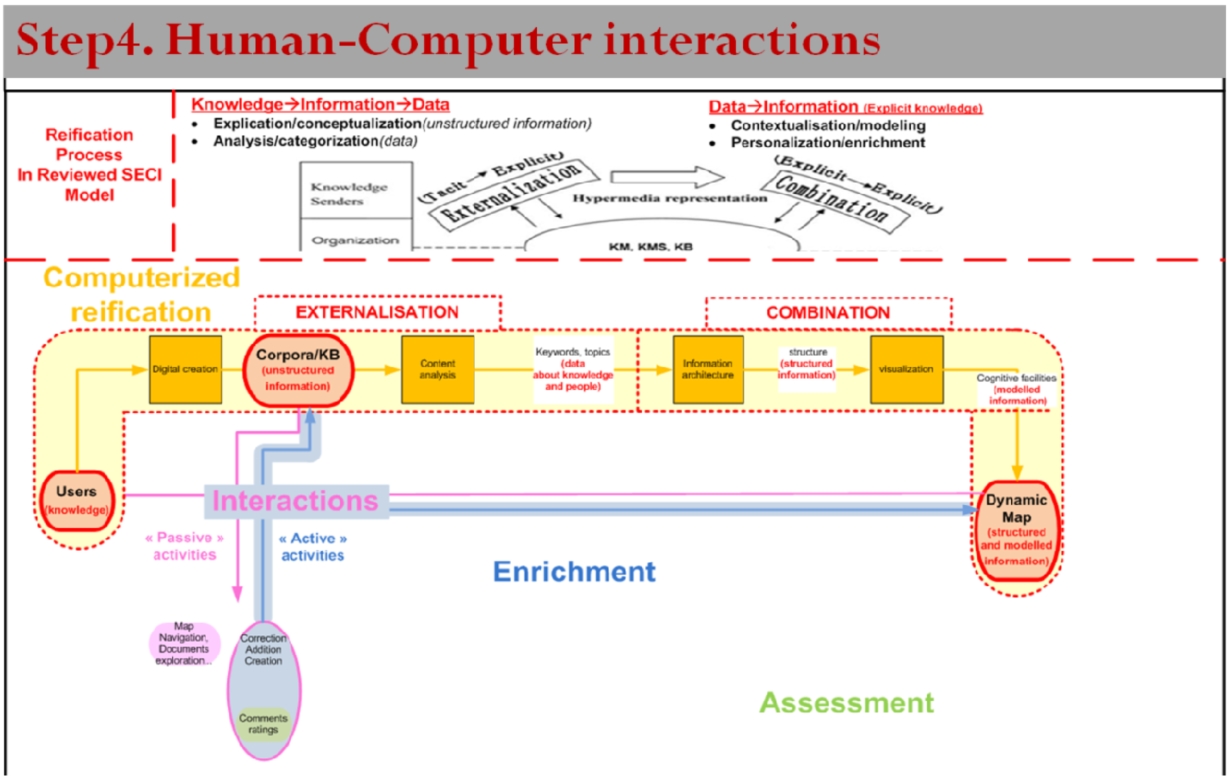
Step 2 focuses on the frame design, the configuration and the gathering of information about corpora and users.



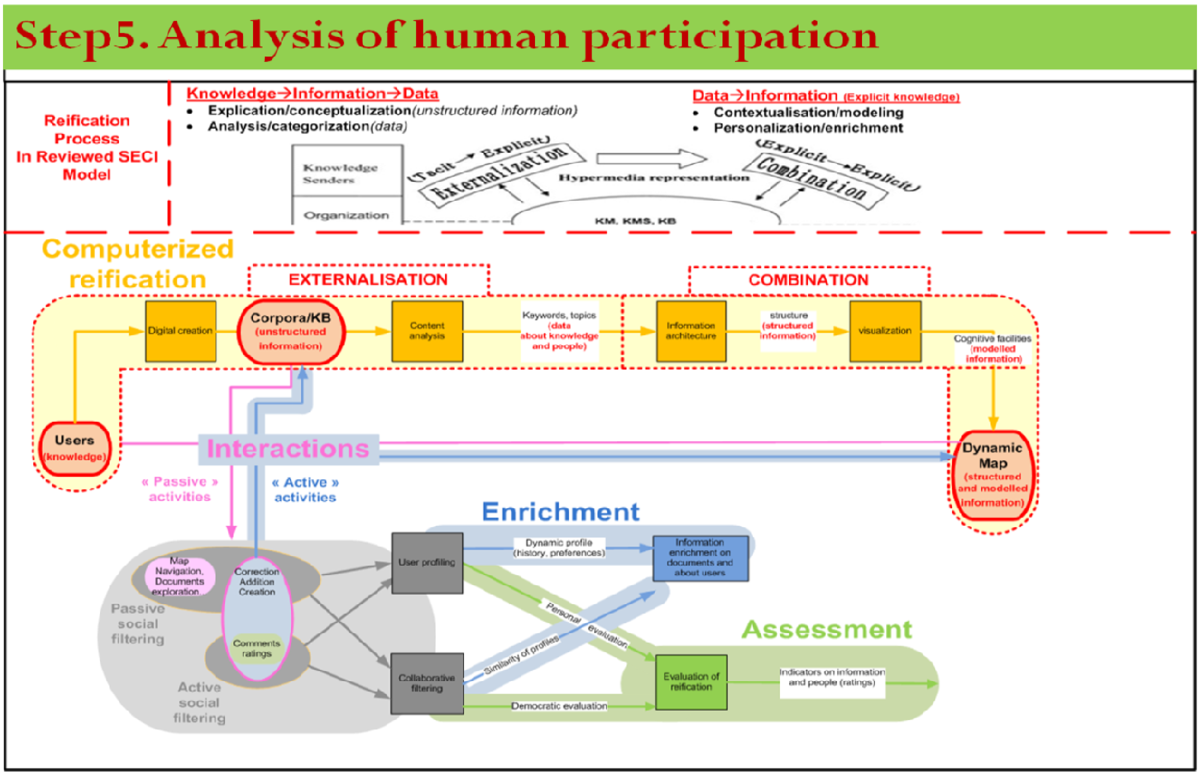
Step 3 analyses corpora in order to help users to emphasize, extract, organize and visualize information contained in knowledge base.



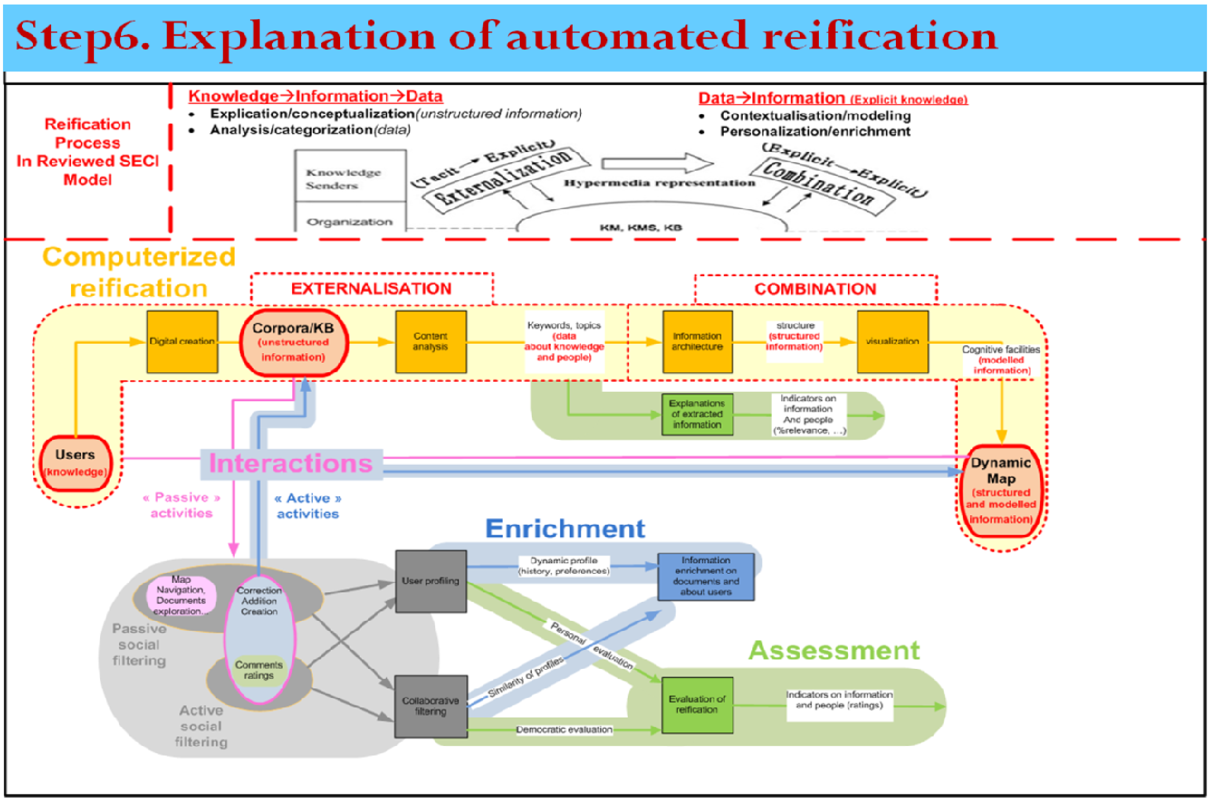
In the step 4 users interact with KB and result of reification (map) in order to control and explain automatic reified information. They can enrich and assess the knowledge base and the dynamic map passively (navigation, exploration, click) or actively (addition, correction, comments, ratings,…).



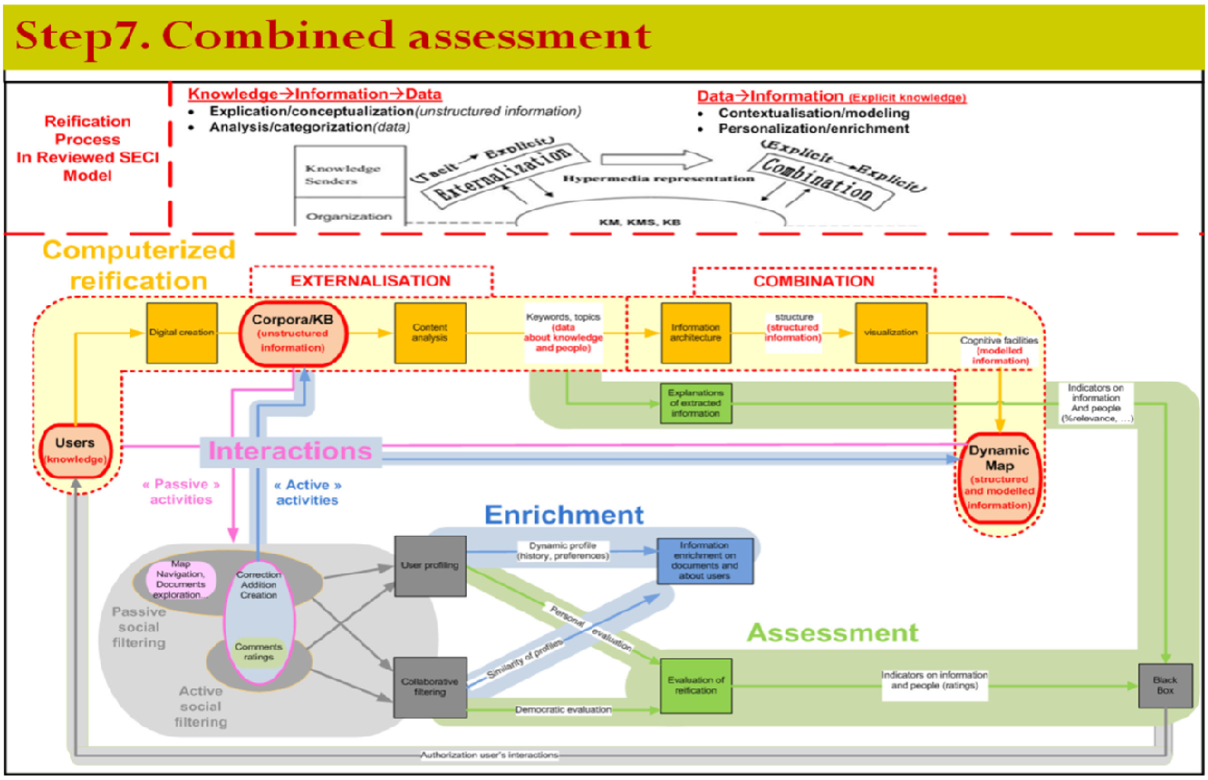
Step 5 is the automatic analysis of human participation, when interactions are used to provide further information, thanks to users profiling and information filtering.



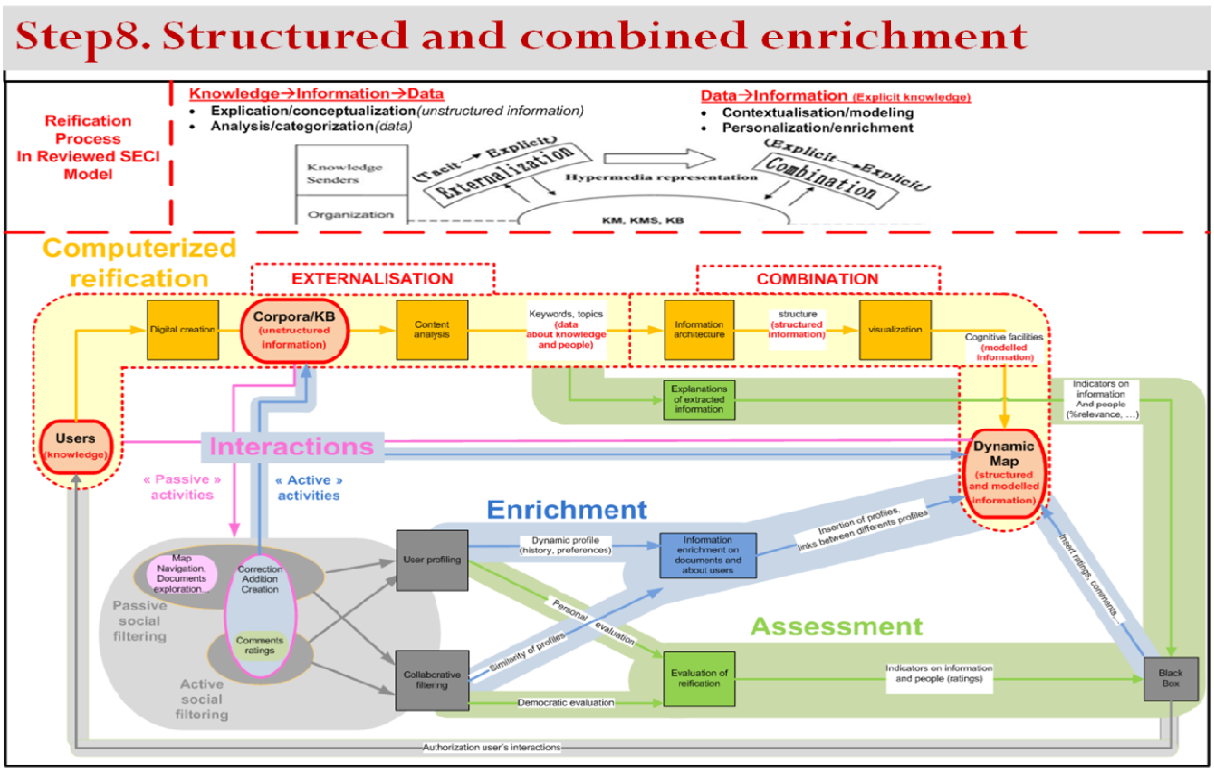
In parallel, automatic extraction can be explained in using the statistics of methods content analysis in step 6. That provides another source of assessment.



In step 7 the different assessment coming from the steps 4, 5 and 6 can be combined, in order to provide good indicators and authorize user's participation.



Finally, the last and 8th step uses combined evaluation and manual (from interactions) and automatic (from information filtering) enrichment to complete map.



APPENDIX 8. RESULTS OF TESTS ON CAT

Corpus Size (S)	Nb of documents (N)	Average document size (L)	Nb of Topics (K)	Time
136738	15	9116	3	95
136738	15	9116	5	83
136738	15	9116	8	100
136738	15	9116	10	337
136738	15	9116	15	391
136738	15	9116	20	464
184839	23	8036	3	180
184839	23	8036	5	269
184839	23	8036	8	630
184839	23	8036	10	653
184839	23	8036	15	395
184839	23	8036	20	1132
287822	57	5050	3	280
287822	57	5050	5	359
287822	57	5050	8	470
287822	57	5050	10	558
287822	57	5050	15	1268
287822	57	5050	20	1313
630430	45	14010	3	1129
630430	45	14010	5	1234
630430	45	14010	8	1370
630430	45	14010	10	2066
630430	45	14010	15	2479
630430	45	14010	20	2496
847051	61	13886	3	2036
847051	61	13886	5	2508
847051	61	13886	8	2961
847051	61	13886	10	2922
847051	61	13886	15	4398
847051	61	13886	20	5356
2928194	145	20194	10	33105

APPENDIX 9. CIRP CASE STUDY: OVERVIEW OF RESULTS

Overview of Results (1)

1. Automatically group papers into categories

GA 2006: Categories 1 to 5				
Category 1	**Category 2**	**Category 3**	**Category 4**	**Category 5**
C05_Ahn.pdf	A06_Bley.pdf	A02_Arai.pdf	C15_min.pdf	A04_Kim.pdf
C07_yamaguchi.pdf	C04_lauwers.pdf	A07_Krueger.pdf	C18_Axinte.pdf	G01_Tricard.pdf
C13_Chandrasekaran.pdf	Dn02_Kimura.pdf	A09_Kara.pdf	Dn01_Miropolsky.pdf	G02_Zhou.pdf
C16_Takeuchi.pdf	Dn07_Kayis.pdf	Dn04_Krause.pdf	E11_Chun.pdf	M02_Zatarain.pdf
E01_Okada.pdf	E07_Zhu.pdf	Dn05_ballu.pdf	E13_Hon.pdf	M04_VanBrussel.pdf
E06_Rombouts.pdf	F15_Bariani.pdf	Dn06_Guttman.pdf	F10_Kleiner.pdf	M05_J_Kim_Corrected.pdf
E08_Park.pdf	G04_Yanagihara.pdf	Dn08_Kim.pdf	F14_Mori.pdf	M06_Sriyotha.pdf
E09_Yu.pdf	O08_Uffmann_Sihn.pdf	M08_Susanu.pdf	G05_Jeong.pdf	M07_Suzuki.pdf
G08_Karpuschewski.pdf	O09_Lucchetta.pdf	M11_Neugebauer.pdf	M09_Erkorkmaz.pdf	M10_Erkorkmaz.pdf
M01_Altintas.pdf	O16_Denkena.pdf	O01 Abele.pdf	M17_Hoshi.pdf	M12_Verl.pdf
M03_Budak.pdf	O17_Zeng.pdf	O04_HEIMaraghy.pdf	S02_Jiang.pdf	M13_Shamoto.pdf
P07_Brinksmeier.pdf		O07_Langaa Jensen.pdf	S06_Bissacco.pdf	P02_Bringmann.pdf
S10_Dambon.pdf		O11_Monostori.pdf	S07_Ohmori.pdf	P05_shore.pdf
		O13_Maropoulos.pdf	S09_Che.pdf	P06_Brecher.pdf
		O20_Kaihara.pdf		P08_Jansen.pdf
		O21_Meier.pdf		P09_Schmitt.pdf
		O22_Butala.pdf		P10_Kuriyama.pdf
				P11_Hidaka.pdf

INDUTECH Your Key to Manage Innovation

GCC Global Competitiveness Centre in Engineering

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (2)

2. Determine descriptive terms for each category

GA 2006: Categories 1 to 5				
Category 1	**Category 2**	**Category 3**	**Category 4**	**Category 5**
tool	product	design	surface	measurement
cutting	process	system	process	axis
machining	assembly	production	high	machine
milling	production	manufacturing	signal	surface
workpiece	model	control	surfaces	control
µm	design	systems	scanning	position
surface	new	network	chemical	system
wear	risk	process	material	mm
mm	planning	time	alloy	error
high	time	management	ball	polishing
material	knowledge	method	data	accuracy
removal	manufacturing	approach	mechanical	positioning
…	…	…	…	…

INDUTECH Your Key to Manage Innovation

GCC Global Competitiveness Centre in Engineering

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (3)

3. Determine overlaps between categories in terms of descriptive terms

Category 1 & 2	Category 1 & 3	Category 1 & 4	Category 1 & 5
	system	system	process
	process	error	
	manufacturing	time	
	planning		
Category 2 & 3	**Category 2 & 4**	**Category 2 & 5**	
system		cutting	
process		tool	
manufacturing		machining	
systems		mm	
Category 3 & 4	**Category 3 & 5**		
design	process		
system			
time			
Category 4 & 5			

INDUTECH Your Key to Manage Innovation — GCC Global Competitiveness Centre in Engineering — UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (4)

4. Determine papers conceptually similar to a given paper

Year	2004	2002 to 2006
STC	Dn	
Filename	Dn01_Lutters.pdf	
Most Similar Document	Dn12_Hon.pdf (2004)	52-1-2003-109.pdf (2003, STC Dn)
2nd Most Similar Document	Dn11_Brissaud.pdf (2004)	A03_Janz.pdf (2006)
3rd Most Similar Document	Dn03_Jin.pdf (2004)	O09_Monitto.pdf (2002)

INDUTECH Your Key to Manage Innovation — GCC Global Competitiveness Centre in Engineering — UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (5)

5. Determine descriptive terms for each paper

Year	STC	Filename	One-word Terms	Two-word Terms
2004	Dn	Dn01_Lutters.pdf	design	information content
			information	driving aspects
			process	workflow management
			content	driven aspects
			management	chain reaction
			product	point view
			processes	task networks
			aspects	processes involved
			task	support system
			design	proc int

INDUTECH Your Key to Manage Innovation

GCC Global Competitiveness Centre in Engineering

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (6)

6. Determine descriptive terms for each STC

Year	STC	One-word Terms	Two-word Terms
All	C	cutting	cutting edge
		tool	tool life
		chip	cutting speed
		surface	finite element
		machining	flank wear
		wear	depth cut
		material	chip thickness
		mm	cutting force
		edge	chip formation
		model	tool path

INDUTECH Your Key to Manage Innovation

GCC Global Competitiveness Centre in Engineering

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

[1] Invalid title extraction due to the lack of a carriage return character between the title and the abstract.

Overview of Results (7)

7. Extract metadata from GA papers

Year	2006			
STC	S			
Filename	S11 Takaya.pdf			
Title	A Novel Surface Finishing Technique for Microparts Using an Optically Controlled Microparticle Tool Abstract This paper focuses on the surface fi nishing of a micropart made of single-crystal silico			
Authors	*Unable to extract*	*Unable to extract*	*Unable to extract*	*Unable to extract*
Abstract	*Unable to extract*			
Listed Keywords	*Unable to extract*			
Proper Nouns	Novel Surface Finishing Technique	Microparts Using	Optically Controlled	Microparticle Tool

INDUTECH Your Key to Manage Innovation — GCC Global Competitiveness Centre in Engineering — UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (8)

8. Determine descriptive terms for each author (cont.)

S. Tichkiewitch	Keywords	Key Terms
	cutting	vibration drilling
	design	design activity
	vibration	integrative environment
	software	spindle speed
	drill	international journal
	drilling	vibratory drilling
	ideve	cutting interruption
	chip	share knowledge
	tool	chip thickness
	environment	limit cycle
	drilling	common language
	dynamics	twist drill
	product	
	svdh	

INDUTECH Your Key to Manage Innovation — GCC Global Competitiveness Centre in Engineering — UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (9)

9. Determine STC where a given paper would fit best

Year	**2006**			
Filename	**M12_Verl.pdf**			
Title	Force Free Add-on Position Measurement Device for the TCP of Parallel Kinematic Manipulators			
Authors	A.Verl	N. Croon	C. Kramer	T. Garber
Original STC	**M**			
Calculated STC	**F**	**M**	**O**	
Abstract	The position of the tool center point (TCP) of a parallel kinematic manipulator (PKM) is traditionally being measured indirectly by means of the position measurement of the drives. Cutting forces and acceleration forces cause displacements of the TCP, which cannot be detected from the position measurement of the drives. To improve the position accuracy of the TCP a force free add-on position measurement device is suggested. The kinematic design of such a measurement device, the calibration and its application for feedback control and improved TCP positioning in the presence of external forces is described. Experimental results are presented to illustrate the expected improvements in TCP positioning.			
Listed Keywords	**Parallel Kinematic Manipulator**	**Accuracy Improvement**	**Measurement Device**	

INDUTECH Your Key to Manage Innovation GCC Global Competitiveness Centre in Engineering UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (10)

10. Determine outlier papers for a given STC

STC	**S**			
Filename	**S07_Lee.pdf**			
Title	**Dynamics and Control of Tapping Tip in Atomic Force Microscope for Surface Measurement Applications**			
Authors	S. I. Lee	J. M. Lee	S. H. Hong	
Year	**2005**			
Abstract	In tapping mode atomic force microscopy (TM-AFM), the vibro-contact response of a resonating tip is used to measure the nanoscale topology and other properties of a sample surface. However, the nonlinear tip-surface interactions can affect the tip response and destabilize the tapping mode control. Especially it is difficult to obtain a good scanned image of high adhesion surfaces such as polymers and biomolecules using conventional tapping mode control. In this study, theoretical and experimental investigations are made on the nonlinear dynamics and control of TM-AFM. Also we report the surface adhesion is an additional important parameter to determine the control stability of TM-AFM. In addition, we proved that it was adequate for the soft and high adhesion sample to be modeled with JKR contact to obtain a reasonable tapping response in AFM.			
Listed Keywords	Atomic force microscopy (AFM)	Tapping mode	Nonlinear dynamics	

INDUTECH Your Key to Manage Innovation GCC Global Competitiveness Centre in Engineering UNIVERSITEIT·STELLENBOSCH·UNIVERSITY jou kennisvennoot·your knowledge partner

Overview of Results (11)

11. Determine most conforming papers for a given STC

STC	**S**			
Filename	**S01_Namba.pdf**			
Title	**Surfaces of Calcium Fluoride Single Crystals Ground with an Ultra-Precision Surface Grinder**			
Authors	Y. Namba	T. Yoshida	S. Yoshida	K. Yoshida
Year	**2005**			
Abstract	Calcium fluoride single crystals for next-generation optical lithography were fabricated with surfaces corresponding to the (001), (111), and (110) crystalline planes. The grinding process utilized an ultra-precision surface grinder and was optimized for resin-bonded SD3000-75-B diamond wheels. A premium crystalline surface had a measured surface roughness of 0.89 nm *Ra,* 6.99 nm *Ry* and 1.10 nm rms, with no microcracks. We show that the surface roughness of these samples depends on the crystalline plane and the orientation, as well as on the grinding conditions. Measurement of the sub-surface damage layer and the laser-induced damage threshold are also presented herewith.			
Listed Keywords	Ultra-precision Grinding	Crystalline Anisotropy	Calcium Fluoride	

References

1. **AFGI, Association Française de Gestion Industrielle. 1992.** Evaluer pour évoluer,les indicateurs de performance au service du pilotage. 1992.

2. **Ahmed, Kal. 2000.** Topic Maps for Repositories, Proceedings of XML Europe 2000. [En ligne] 2000. http://www.gca.org/papers/xmleurope2000/papers/s29-04.html.

3. **Amende, Nadine et Groschupf, Stefan. 2004.** *Visualizing an Auto-Generated Topic Map.* Halle-Wittenberg : s.n., 2004.

4. **American Library Association. 1999 .** Task Force on Metadata Summary Report. 1999 .

5. **Ammar-Khodja, S.,. 2006.** *La gestion des connaissances, principe et méthodes : Course of Knowledge Management in Centrale Nantes.* 2006.

6. **Archer, N. 2006.** A classification of communities of practices. *In Idea Group – Encyclopedia Of Communities Of Practice In Information And Knowledge Management.* 2006.

7. **Associate members of the VRL. 2004.** *The overview for the Associate members of the VRL – The VRL Joint Programme of Activities, 2nd stage version .* 2004.

8. **Bachimont, B. 1996.** *Herméneutique matérielle et Artéfacture : des machines qui pensent aux machines qui donnent à penser ; Critique du formalisme en intelligence artificielle. Thèse de doctorat d'épistémologie.* s.l. : École Polytechnique, 1996.

9. **Baroni de Carvalho, R et Araujo Tavares Ferreira, M. 2002.** Using information technology in support knowledge conversion processes. *Information Research, Vol. 7 No. 2.* January 2002.

10. **Berelson, B. 1952.** *Content Analysis in Communication Research.* Glencoe, Ill : Free Press, 1952.

11. **Berrah, L. 2002.** *Les indicateurs de performance : concepts et applications .* s.l. : Cépadues, 2002.

12. **Bescos, P.L., et al. 1995.** *Contrôle de gestion de Management.* s.l. : Editions Montchestien, Collection Entreprendre, Guide des techniques et de la décision, 1995.

13. **Block, Sharon. 2006.** Tales from the Vault : Doing More with Digitization. *Common Place.* [Online] January 2006. http://www.common-place.org/vol-06/no-02/tales/.

14. **Brown, J. S. and Duguid, P. 1991.** Organizational learning and communities-of-practice: Toward a unified view of working, learning and innovation. *Organization Science n°2.* 1991, pp. 40-57.

15. **Brown, J.S et Duguid, P. 2000.** *The social life of information.* Boston, MA : Harvard Business School Press, 2000.

16. **Bultermann, D. C. A. 2004.** Is It Time For a Moratorium on Metadata? *IEEE MultiMedia,.* 2004.

17. **Burke, Robin. 2000.** Knowledge-based recommender systems. *Encyclopedia of Library and Information Science.* University of California, Irvine : s.n., 2000.

18. **Canard, F. et Carton, S. 2004.** *Apport des référentiels qualité à une meilleure capitalisation des connaissances : la qualité liée aux applications informatiques, Etude de la réglementation 21 CFR Part 11.* 2004.

19. **Candlot, A. 2006.** *IVGI Team Benchmarking on what is Knowledge!* Nantes : s.n., 2006.

20. **Card, Stuart K. 1999.** *Reading in information visualization: using vision to think.* [éd.] Jock D. Mackinlay et Ben Schneiderman. San Francisco : Morgan Kaufmann, 1999.

21. **Charlet, J. 2002.** *L'ingénierie des connaissances, développement, résultats et perspectives pour la gestion des connaissances médicales : Mémoire d'habilitation à diriger des recherches.* 2002.

22. **Chauve, Maxime. 2007.** *Bibliographical synthesis: Performance Evaluation of Supply Network.* Bath : s.n., 2007.

23. **Chen, Chaomei. 2003.** *Mapping scientific frontiers: The Quest for Knowledge Visualization.* London : Springer-verlag, 2003.

24. **cineca. 2006.** www.cineca.it/gai/files/pdf/Mole_en.pdf. [En ligne] 2006.

25. **contemplativemind. 2007.** http://www.contemplativemind.org. [En ligne] August 2007 2007.

26. **Croassdell, D.T, et al. 2003.** A meta-analysis of methodologies for research in knowledge management, organizational learning and organizational memory. *Five years at HICSS, System Sciences, Proceedings of the 36th Annual Hawaii International Conference, on System Sciences (HICSS'03).* 2003, 4.

27. **Delone, W.H. et Lean, E.R. Mc. 1992.** Information System Success: the quest for the dependent variable. *Information System Research 3.* 1992, pp. 60-95.

28. **Dodge, Martin. 2000.** NewsMaps: Topographic Mapping of information. *http://mappa.mundi.net/maps/maps_015/#ref_2.* [En ligne] 2000.

29. **Du Preez, Nick. 2004.** *The associate partners presentation in Grenoble.* 2004.

30. **Dürsteler, Juan C. 2004.** Conceptual Maps. *Inf@Vis! The digital magazine of InfoVis.net.* [Online] 03 01, 2004. http://www.infovis.net/printMag.php?num=141&lang=2.

31. **Edinburgh University. 2007.** http://eweb.chemeng.ed.ac.uk/courses/control/restricted/course/second/course/lecture1.html. [En ligne] 2007. [Citation : 16 August 2007.]

32. **Ermine, J.-L. 1996.** *Les systèmes de connaissance.* s.l. : Hermès, 1996.

33. **Even-Zohar, Yair. 2002.** Introduction to Text Mining. *algdocs.ncsa.uiuc.edu/PR-20021116-2.ppt .* [En ligne] 2002.

34. **Feldman, R and Dagan, I. 1995.** *Kdt - knowledge discovery in texts. In Proc.of the First Int. Conf. on Knowledge Discovery (KDD).* 1995. pp. 112–117.

35. **Giraudel, J. L. et Lek, S. 2001.** *A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination.* 2001.

36. **Goecks, Jeremy and Shavlik, Jude. 2000.** *Learning Users' Interests by Unobtrusively Observing their normal behavior.* Madison : s.n., 2000.

37. **Goldberg, David, et al. 1992.** Using collaborative filtering to weave an information tapestry. *Communication of the ACM.* 1992, Vol. 35, pp. 61-70.

38. **GoodwinJones, R. 2003.** *Emerging technologies. Blogs and wikis: environments for online Collaboration. Language Learning & Technology.* s.l. : Michigan State University, Department of Language Learning and Technology, 2003.

39. **Greenwood, R, et al. 2006.** *Knowledge Mapping and Bringing about Change for the Sustainable Urban Environment.* London : EPSRC, 2006.

40. **Grokker. 2007.** http://www.grokker.com/. [En ligne] 2007. [Citation : 10 July 2007.]

41. **Gruber, T.R. 1995.** Towards principles for the design of ontologies used for knowledge sharing. *International Journal on Human-Computer Studies.* 1995, pp. 907-928.

42. **HCIL. 2007.** Human Computer Interaction Lab (University of Maryland). *http://www.cs.umd.edu/hcil/research/visualization.shtml.* [En ligne] 2007. [Citation : 10 July 2007.]

43. **Herlocker, J., Konstan, J. et and Riedl, J. 2000.** Explaining Collaborative Filtering Recommendations. *Proceedings of the ACM 2000 Conference on Computer Supported Cooperative Work.* 2000.

44. **Hertzum, M et Petjersen, A.M. 2000.** Information Seeking Practices of Engineers: Searching for Documents as well as for People. *Information Processing and Management 36.* 2000.

45. **Hildreth, P et Kimble, C. 2006.** *Knowledge Networks: Innovation through Communities of Practice.* s.l. : Idea Group Publishing, 2006.

46. **Holsti, O.R. 1969.** *Content Analysis for the Social Sciences and Humanities.* Reading, MA : Reading, MA, 1969.

47. **Huang, Hsiu-Mei et Liaw, Shu-Sheng. 2004.** *The Framework of Knowledge Creation for Online Learning Environments.* s.l. : Canadian Journal of Learning and Technology, 2004. Vol. 30.

48. **Indutech. 2007.** Indutech : Your Key to Manage Innovation. [Online] Indutech (pty) Ltd, 2007. http://www.indutech.co.za.

49. **Iribarne, P. 2006.** *Les tableaux de bord de la performance.* s.l. : Dunod, 2006.

50. **ISO/IEC13250. 1999.** *Topic Maps.* 1999.

51. **ISO/IEC9126. 2001.** Software Quality. 2001.

52. **J.-P., Poitou. 1996.** La gestion des connaissances comme condition et résultat de l'activité industrielle. *Intellectica.* 1996.

53. **Jaime, A. 2005.** *From Quality Management to Knowledge Management in Research Projects : An Approach through the Management of Contents in Bibliographical Research .* 2005.

54. **Janis, I.L. 1949.** The problem of validating content analysis. [book auth.] Nathan Leites, & Associates Harold D. Lasswell. *Language of politics: Studies in quantitative semantics.* New York : George W. Stewart, Publisher, Inc., 1949, pp. 55-82.

55. **Judelman, G.B. 2004.** *Knowledge Visualization: Problems and Principles for mapping the Knowledge Space.* 2004.

56. **Kimble, C, Hildreth, P and Wright, P. 2000.** Communities of Practice: Going Virtual. *Knowledge Management and Business Model Innovation, Idea Group Publishing.* London/Hershey : s.n., 2000.

57. **Kohonen, T. 2001.** Self-Organizing Maps. Berlin : s.n., 2001.

58. **Koike, H. 1995.** Fractal views: a fractal-based method for controlling information display. *ACM Transactions on Information Systems.* 1995, pp. 305-323.

59. **Krippendorff, K. 1980.** *Content Analysis: An Introduction to Its Methodology.* Newbury Park, CA : Sage Publications, 1980.

60. **Krippendorff, Klaus. 1986.** *A Dictionary of Cybernetics.* s.l. : unpublished, 1986.

61. **Lasswell, Harold. 1948.** The structure and function of communication in society. *The Communication of Ideas.* Lyman Bryson. New York : Harper & Brothers, 1948.

62. **Lave, J et Wenger, E. 1991.** *Situatd Learning. Legitimate Peripheral Participation Cambridge University Press.* 1991.

63. **Le Grand, B. et Soto, M. 2003.** Topic Maps Visualization. [éd.] V. Geroimenko et C. Chen. *Visualizing the Semantic Web. XML-based Internet and Information Visualization.* 2003.

64. **Lebas, M. 1995.** « Oui, il faut définir la performance ». *Revue française de comptabilité n°269 .* 1995.

65. **Lemeur, Loic. 2007.** http://www.loiclemeur.com/IPSOSeuroblogs2006.pdf. [En ligne] 2007. [Citation : 21 08 2007.]

66. **Lesser, E.L. and Storck, J. 2001.** Communities of practice and organizational performance. *IBM Systems Journal n°40.* 2001, pp. 831-841.

67. **Li, Charlene. 2007.** *Social Technographics®, Mapping Participation In Activities Forms The Foundation Of A Social Strategy.* s.l. : Forrester, 2007.

68. **Malone, et al. 1987.** systems, Intelligent Information-sharing. *Communications of the ACM.* 1987, Vol. 30, pp. 390-402.

69. **Maltz, David et Ehrlich, Kate. 1995.** Pointing the way: active collaborative filtering. *http://acm.org/sigchi/chi95/proceedings/papers/ke_bdy.htm.* [En ligne] 1995.

70. **Marti, Hearst. 1999.** *Untangling text data mining. In Proceedings of ACL '99: The 37th annual meeting of the Association for Computational Linguistics.* University of Maryland : s.n., 1999. Also available at http://www.sims.berkeley.edu/~hearst/papers/acl99/acl99-tdm.htm.

71. **McKeone, D. 1995.** *Measuring Your Media Profile.* London : Gower, 1995.

72. **Mons, Alexis. 2006.** http://www.groupereflect.net/blog/archives/2006/05/la_regle_du_1.html#more. [En ligne] 2006.

73. **Moor, Aldo de et Smits, Martin. 2002.** *Key Performance Indicators for Knowledge Management in a community of practices.* Tilburg : Tilburg University, 2002.

74. **Netvis. 2007.** http://www.netvis.org/. *Netvis.* [En ligne] Netvis, 2007. [Citation : 16 August 2007.]

75. **Nonaka, I, Toyama, R et Konno, N. 2000.** SECI, Ba and Leadership: A unified Model of Dynamic knowledge creation. *Long Range Planning.* february 2000, Vol. 33, 1, pp. 5-34.

76. **Novak, J.D and Canas, A.J. 2006.** *The Theory Underlying Concept Maps and How to Construct Them, Technical Report IHMC CmapTools 2006-01.* s.l. : Florida Institute for Human and Machine Cognition, 2006.

77. **O'Reilly, T. 2005.** *from http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html .* 2005.

78. **Oliveira, Rocha Alexandre. 2000.** Deployment of Topic Maps for Navigation and Searching in Huge Information Spaces as Component of Learning Environment. 2000.

79. **Palmer, Shelly. 2006.** *Television Disrupted: The Transition from Network to Networked TV.* s.l. : Focal Press, 2006. p. 200.

80. **Petersen, Kenneth J., Monc, Robert M and Ragatz, Gary L. 2005.** an examination of collaborative planning effectiveness and supply chain performance. *the journal of supply chain management.* 2005, pp. 14-25.

81. **Poitou, J.-P. 1996.** *La gestion des connaissances comme condition et résultat de l'activité.* 1996.

82. **Polanyi, M. 1966.** *The tacit dimension.* New York : Anchor Day, 1966.

83. **Prusak, L and Lesser, E.L. 1999.** Communities of practice, social capital and organizational knowledge. *Information Systems Review.* 1999, pp. 3-9.

84. **Rauffet, P. 2007.** *Mémoire Bibliographique: Etat de l'Art sur le Knowledge Management.* Nantes : s.n., 2007.

85. **Rigouste, L, Cappé, O et Yvon, F. 2006.** *Quelques observations sur le modèle LDA.* JADT 2006 : 8es Journées internationales d'Analyse statistique des Données Textuelles. 2006.

86. **Rohrer, R. M. et Swing, E. 1997.** Web-Based Information Visualization. *IEEE Computer Graphics and Application.* 1997.

87. **Ryszard Kruk, S et Decker, S. 2005.** *Making Social Collaborative Filtering Real.* Galway : Digital Enterprise Research Institute, 2005.

88. **Schoen, S. 2001.** *Gestaltung und unterstützung von communities of practice.* München : Herbert Utz Verlag, 2001.

89. **Sekkat, S et Canonne, R. 2005.** *L'apport des Systèmes à Base de Connaissance en gestion de production.* Casablanca : CPI'2005, 2005.

90. **Senechal, O. 2004.** *Mémoire d'habilitation à diriger des recherches: Pilotage des systèmes de production vers la performance globale.* Université de Valenciennes et du Hainaut Cambresis : s.n., 2004.

91. **Simoudis, E. 1996.** *Reality check for data mining. In IEEE Expert, 11(5).* 1996.

92. **The Hive Group. 2007.** www.hivegroup.com. [En ligne] 2007. [Citation : 11 July 2007.]

93. **Theard, J.B. 2005.** Ontology for change management projects : perspectives, methods,impacts,… on change projects: report of internship in Indutech. 2005.

94. **TouchGraph. 2007.** http://touchgraph.com. [En ligne] 2007. [Citation : 5 June 2007.]

95. **Tsai, W and Goshal, S. 1998.** Social capital and value creation: The role of intrafirm networks. *Academy of Management Journal n°41.* 1998, pp. 464-476.

96. **UCI. 2006.** UCIrvine : Donald Bren School of Information and Computer Science. [Online] July 2006. http://www.ics.uci.edu/community/news/press/view_press?id=51.

97. **US General Accounting Office. 1996.** *Content Analysis: A Methodology for Structuring and Analyzing Written Material.* Washington, D.C. : s.n., 1996.

98. **Uys, Wilhelm. 2007.** Improved Utilisation of Organisational Documents Using a Conceptual Framework. Indutech, 2007.

99. **Vatant, B. 2003.** *Ontology-driven topic maps.* 2003.

100. **VRL-KCiP. 2006.** *Task 105 - Deliverable No. [105] Specification of the services of the VRL-KCiP Knowledge Management System.* Stockholm : s.n., 2006.

101. **—. 2007.** VRL Knowledge Community in Production. [Online] VRL-KCiP, 2007. http://www.vrl-kcip.org/.

102. **W3C. 2004.** W3C : Web-Ontology (WebOnt) Working Group . [Online] 2004. http://www.w3.org/2001/sw/WebOnt/.

103. **Weber, R.P. 1990.** *Basic Content Analysis.* 2nd. Newbury Park, CA : Sage Publications, 1990.

104. **Wenger, E. 1998.** *Communities of Practice: Learning, Meaning and Identity.* s.l. : CUP, 1998.

105. **Wenger, E, McDermott, R and Snyder, W.M. 2002.** *Cultivating communities of practice.* Boston : Harvard Business School Press, 2002.

106. **Werthner, H, et al. 2003.** Harmonise - towards Interoperability in Tourism Domain. s.l. : ECT RL, 2003.

107. **Yiman Seid, D et Kobsa, A. 2003.** *Expert-finding organizations : Problem and Domain analysis and the DEMOIR approach, Sharing expertise : Beyond Knowledge Management.* s.l. : The MIT Press, 2003.

108. **Young, Julian. 1992.** *Nietzsche's Philosophy of Art* . Cambridge : Cambridge University Press, 1992.

109. **Zimmermann, A, Lorenz, A and Specht, M. 2002.** Reasoning From Contexts. *Personalization for the Mobile World: Workshop Proceedings on Adaptivity and User Modeling in Interactive Systems (ABIS).* Hannover : Henze,N., 2002, pp. 114-120.